Signal Analysis
Representation Tools

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische
Universiteit Eindhoven, op gezag van de Rector Magnificus,
prof.dr. M. Rem, voor een commissie aangewezen door het
College voor Promoties in het openbaar te verdedigen op
donderdag 15 februari 2001 om 16.00 uur

door

Bahaa Eddine Sarroukh

geboren te Larache
Dit proefschrift is goedgekeurd door de promotoren:

prof.dr.ir. M.L.J. Hautus
en
prof.dr.ir. J.W.M. Bergmans

Copromotor:

dr.ir. A.C. den Brinker

©Copyright 2001 Bahaa Eddine Sarroukh
All rights reserved. No part of this publication may be reproduced, stored in a re-
trieval system, or transmitted, in any form or by any means, electronic, mechanical,
photocopying, recording or otherwise, without the prior written permission from the
copyright owner.

CIP-DATA LIBRARY TECHNISCHE UNIVERSITEIT EINDHOVEN
Sarroukh, Bahaa Eddine

Signal Analysis: representation tools / by Bahaa E. Sarroukh. –
NUGI 832
Trefw.: orthogonale reeksen / wiskundige filters / wiskundige transformaties
/ signaalverwerking.
Subject headings: approximation theory / time-frequency analysis /
time-domain analysis / signal processing.
Promotiecommissie

prof.dr.ir. J.W.M. Bergmans  
Department of electrical engineering, TU/e Eindhoven

prof.dr.ir. O.J. Boxma  
Department of mathematics and computing science, TU/e Eindhoven

dr.ir. A.C. den Brinker  
Philips research, Eindhoven

dr.ir. S.J.L. van Eijndhoven  
Department of mathematics and computing science, TU/e Eindhoven

prof.dr.ir. M.L.J. Hautus  
Department of mathematics and computing science, TU/e Eindhoven

prof.dr.ir. A.A.C.M. Kalker  
Department of electrical engineering, TU/e Eindhoven

prof.dr. J.K. Lenstra  
Department of mathematics and computing science, TU/e Eindhoven

prof.dr.ir. P.M.J. Van den Hof  
Department of applied physics, TU Delft

dr.ir. R.N.J. Veldhuis  
IPO center for user-system interaction, TU/e Eindhoven
Summary

Signal representations play a fundamental role in signal processing and in particular in the design of signal processing systems. In fact, the efficiency and the quality of a signal processing system depend strongly on the chosen representation of the signals. Typically, accurate approximations and fast algorithms are needed to represent or analyze signals. In this thesis, two types of signal representations are considered:

1. *parameterized signal representations involving the entire signal domain*;
2. *local representations induced by short-term transformations*.

The second type of signal representation characterizes local properties of signals in relationship with the signal domain, whereas the first type characterizes their global properties.

In this thesis, a variety of problems concerning efficiency and quality is tackled, and corresponding results are derived.

An enforced convergence criterion for well-known series expansions (Jacobi, Hermite, Laguerre, Kautz) is established and, relative to this criterion, optimal parameter choices for these expansions are derived.

The concept of the windowed unitary transformation is introduced, generalizing the concept of short-term Fourier transformation. The links between the concept of windowed unitary transformation and the concept of filter bank are discussed. Easily interpretable representation schemes for perfect reconstruction filter banks are developed.

Boundary artifacts caused by straightforward signal segmentation due to the use of rectangular windows can be reduced replacing the rectangular window by a smooth window and, additionally, by using window symmetry at the boundaries. The corresponding local overlapping signal representations are described for one- and two-dimensional signal domains.
Contents

Summary v

List of Symbols xi

1 Scope of the thesis 1
  1.1 Global representations ............................................. 1
  1.2 Local representations ............................................. 4
  1.3 Applications .......................................................... 6
  1.4 Goals of the thesis ................................................... 8
    1.4.1 Analysis and design by global series representations ...... 8
    1.4.2 Filter banks in a design perspective ......................... 9
    1.4.3 Analysis by domain segmentations ........................... 9
  1.5 Outline of the thesis ............................................... 9
  1.6 Main results of the thesis ......................................... 11

I Analysis and design by global series representations 13

2 Introduction to Part I 15
  2.1 The scope of Part I .................................................. 17
  2.2 Optimal parameters ................................................. 17
  2.3 Analysis by Hermite transformations ............................ 19
  2.4 Design of causal systems .......................................... 20

3 Optimal parameters in orthonormal series expansions 23
  3.1 Optimality condition ................................................ 24
    3.1.1 Continuous-time case ........................................... 27
    3.1.2 Discrete-time case ............................................. 29
  3.2 Second-order operator ............................................. 30
    3.2.1 Second-order differential operator ........................... 31
    3.2.2 Second-order difference operator ............................ 33
  3.3 Optimal parameters in Jacobi expansions ....................... 35
  3.4 Charlier and Meixner optimal parameters ....................... 40
### Contents

3.5 Discussion ......................................................... 43

4 Signal analysis by modulated Hermite transformations .......... 45
  4.1 The considered class of signals .......................... 45
  4.2 The modulated Hermite functions ....................... 46
  4.3 Properties .................................................. 47
  4.4 Compaction criterion ................................. 49
  4.5 Optimum compaction parameters .................... 50
  4.6 Discussion ................................................. 52

5 Design of causal systems by Kautz transformations .......... 55
  5.1 Optimal parameters in modulated Laguerre expansions .... 56
    5.1.1 Definitions and properties ........................ 56
    5.1.2 Criterion and results ............................. 60
  5.2 Optimal parameters in Kautz series expansions .......... 62
    5.2.1 Kautz and Laguerre series ........................ 63
    5.2.2 Optimal discrete Laguerre parameter ................ 67
    5.2.3 Optimal Kautz parameter .......................... 71
    5.2.4 Repeated complex conjugated pole pair .............. 72
    5.2.5 Recursive solution for optimal Kautz parameters ... 75
    5.2.6 Highly oscillating systems ........................ 82
  5.3 Discussion ................................................. 82

II Filter banks in a design perspective ......................... 85

6 Introduction to Part II ........................................ 87
  6.1 Uniformly downsampled filter bank ...................... 88
  6.2 Non-uniformly downsampled filter bank .................. 90
  6.3 Goals ...................................................... 91
  6.4 Outline ................................................... 91

7 Uniformly downsampled filter banks .......................... 93
  7.1 Basic building blocks in filter banks ................... 94
  7.2 Structures and perfect reconstruction in filter banks .... 99
  7.3 Equivalence between filter banks and frames ............ 106
  7.4 Localizing unitary transformations and filter banks .... 114
  7.5 Linear constant coefficient difference equations ......... 121
    7.5.1 Perfect reconstruction in the critical case ......... 128
    7.5.2 Perfect reconstruction in the oversampled case ....... 129
  7.6 Kautz and Laguerre filter banks ........................ 131
List of Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathbb{N}, \mathbb{N}_0 )</td>
<td>the sets of positive integers and non-negative integers, respectively</td>
</tr>
<tr>
<td>( \mathbb{Z}, \mathbb{Z}^+, \mathbb{Z}^- )</td>
<td>the sets of integers, positive integers and negative integers, respectively</td>
</tr>
<tr>
<td>( \mathbb{R} )</td>
<td>the set of real numbers</td>
</tr>
<tr>
<td>( \mathbb{C} )</td>
<td>the set of complex numbers</td>
</tr>
<tr>
<td>( \mathbb{C}^+, \mathbb{C}^- )</td>
<td>the right and left half-planes of the complex plane, respectively</td>
</tr>
<tr>
<td>( \mathbb{U} )</td>
<td>the unit circle</td>
</tr>
<tr>
<td>( \mathbb{D} )</td>
<td>the unit disc ( \mathbb{D} := { z \in \mathbb{C}</td>
</tr>
<tr>
<td>( \mathbb{E} )</td>
<td>the exterior of ( \mathbb{D} \cup \mathbb{U}, \mathbb{E} = { z \in \mathbb{C}</td>
</tr>
<tr>
<td>( \forall^* )</td>
<td>the zero is excluded from ( \forall )</td>
</tr>
<tr>
<td># ( \forall )</td>
<td>the number of elements of ( \forall )</td>
</tr>
<tr>
<td>( \text{span} \forall )</td>
<td>the space of all linear combinations of ( \forall )</td>
</tr>
<tr>
<td>( \forall \setminus \mathbb{W} )</td>
<td>the set of elements in ( u \in \forall ), such that ( u \notin \mathbb{W} )</td>
</tr>
<tr>
<td>( \forall^\perp )</td>
<td>denotes the orthogonal complement of ( \forall )</td>
</tr>
<tr>
<td>( \theta )</td>
<td>parameter (scalar or vector)</td>
</tr>
<tr>
<td>( \Theta )</td>
<td>the domain of parameter ( \theta )</td>
</tr>
<tr>
<td>( \mathbb{C}[z] )</td>
<td>the set of polynomials</td>
</tr>
<tr>
<td>( \mathbb{Q}(z) )</td>
<td>the set of rational functions</td>
</tr>
<tr>
<td>( \mathcal{H} )</td>
<td>Hilbert space of signals</td>
</tr>
<tr>
<td>( \mathcal{D} )</td>
<td>set of admissible signals in ( \mathcal{H} )</td>
</tr>
<tr>
<td>( \mathbb{B}(\forall, \mathbb{W}) )</td>
<td>the set of bounded operators with domain ( \forall ) and range in ( \mathbb{W} )</td>
</tr>
<tr>
<td>( \ell_2(\forall) )</td>
<td>the Hilbert space of square-summable functions on ( \forall ), the set ( \forall ) may be of finite length, then ( \ell_2(\forall) ) is the set of functions in ( \ell_2(\mathbb{Z}) ) with support ( \forall \subseteq \mathbb{Z} )</td>
</tr>
<tr>
<td>( \ell_p(\forall) )</td>
<td>Banach space, where ( 1 \leq p &lt; \infty ) specifies the norm in ( \ell_p(\forall) )</td>
</tr>
<tr>
<td>( \mathbb{L}_2(\forall) )</td>
<td>the Hilbert space of square integrable functions on ( \forall )</td>
</tr>
<tr>
<td>( \mathcal{H}_2(\mathbb{E}) )</td>
<td>the Hardy space of square integrable functions on ( \mathbb{U} ) and analytic in the region ( \mathbb{E} ), see Chapter 5 for the definition.</td>
</tr>
<tr>
<td>( \langle x, y \rangle )</td>
<td>denotes the inner product of ( x ) and ( y ) in the Hilbert space where they belong to. Sometimes, we write ( \langle ax, by \rangle = \langle b^* ax, y \rangle ), if the notation</td>
</tr>
</tbody>
</table>
in this way is simplified, even if $a^*x + y$ is not an element of the Hilbert space where the inner product is taken.

- $\|x\|$ the norm of $x$ in the normed space where $x$ belongs to
- $x * y$ denotes the convolution of $x$ and $y$
- $\mathcal{F}$ the Fourier transformation
- $\mathcal{F}_D$ the discrete-time Fourier transformation
- $\mathcal{L}$ the Laplace transformation
- $\mathcal{Z}$ the z-transformation
- $\mathcal{T}$ the left-shift operator
- $\sigma$ the right-shift operator. We note that this operator is defined for the first time in the Part II. In Part I we use the symbol $\sigma$ as a component in a parameter $\theta$.

- $\mathcal{A}[\theta]$ a parameter dependent difference or differential operator
- $\mathcal{A}$ the frame operator
- $\mathcal{J}$ the analysis operator, defined by $\mathcal{J} = \mathcal{A}^* \mathcal{A}$
- $\mathcal{R}_\alpha$ the two-dimensional rotation counterclockwise over the angle $\alpha$
- $\mathcal{P}_{[V, \epsilon, s]}$ an orthogonal projection associated with $V$, dependent of the (scalar or vector) parameters $\epsilon$ and $s$, see Chapters 10 and 11
- $\mathcal{K}^*$ the adjoint operator of the operator $\mathcal{K}$
- $\ker(\mathcal{K})$ the null space or kernel of $\mathcal{K}$
- $\text{range}(\mathcal{K})$ the range of $\mathcal{K}$
- $\text{trace}(\mathcal{E})$ the trace of the matrix $\mathcal{E}$, that is, the sum of the diagonal entries
- $\|\mathcal{K}\|$ the norm of the operator $\mathcal{K}$
- $\mathcal{W}$ a windowed unitary transformation, see Section 7.4 in Chapter 7
- $\mathcal{M}_{i,j}(x)$ represents a measurement on $x$
- $\mathcal{I}$ the identity operator
- $\mathcal{I}_N, I_N$ the $N \times N$ identity matrix
- $\text{support}(x)$ defined by: $t \in \text{support}(x) : \Leftrightarrow x(t) \neq 0$
- $\phi[\theta]$ a function or an operator that depends on the parameter $\theta$
- $\phi[\theta; k]$ a function or a sequence that depends on the parameters $\theta$ and $k$
- $\phi(t; \theta)$ a function acting on $t$ and depending on the parameter $\theta$
- $Q(x; \theta)$ an operator acting on $x$ and depending on the parameter $\theta$
- $B[\theta]$ a collection that depends on the parameter $\theta$
- $B(\theta; k)$ a collection that depends on the parameter $\theta$ and index $k$
- $B$ a library collection of collections of the type $B[\theta]$
- $w[V, \theta]$ $w$ is a window that depends on the set $V$ and has $\theta$ as parameter
- $B^\top$ vector function
- $B^\top$ transpose of $B$
- $z^*$ complex-conjugate of $z$
List of Symbols

\[ \delta \quad \text{Dirac functional} \]
\[ \delta_{mn} \quad \text{Kronecker delta symbol, } \delta_{mn} = 0 \text{ if } m \neq n \text{ and } \delta_{mn} = 1 \]
\[ \binom{m}{n} \quad \text{the binomial coefficient } m(m-1) \ldots (m-n+1)/n! \]
\[ \Gamma(\alpha) \quad \text{the Gamma function} \]
\[ (\alpha)_m \quad \text{Pochhammer symbol, } (\alpha)_m = \alpha(\alpha + 1) \ldots (\alpha + m - 1) \]
\[ \nabla G \quad \text{gradient of } G \]
\[ \Im(z) \quad \text{imaginary part of } z \]
\[ \Re(z) \quad \text{real part of } z \]
\[ \mathbf{1} \quad \text{denotes } (1, 1, 1, \ldots) \text{ in } \ell_1(\mathbb{N}_0) \]
\[ \hat{x} \quad \text{the time reversed of } x \text{ defined by, } \hat{x} = x(-t) \]
\[ \Gamma \quad \text{a grid or a set of indices in } \mathbb{Z}^d, d \in \mathbb{N}_0 \]
\[ [a, b] \quad \text{an interval, in discrete case defined as } \{ t \in \mathbb{Z} | a \leq t \leq b \} \text{ and in continuous case as } \{ t \in \mathbb{R} | a \leq t \leq b \} \]
\[ \oplus \quad \text{denotes the direct sum} \]
\[ \chi_{\mathcal{V}} \quad \text{the characteristic function over } \mathcal{V}, \text{ we denote } \chi_{\mathcal{V}} \mathcal{I} \text{ by } \chi_{\mathcal{V}} \]
Chapter 1

Scope of the thesis

Signal representations play a fundamental role in signal processing and in particular in the design of signal processing systems. In fact, the efficiency and the quality of a signal processing system depend strongly on the chosen representation of the signals. Applications of the systems that we have in mind are in the fields of audio, speech and image processing, communication, system identification and control. Typically, accurate approximations and fast algorithms are needed to represent or analyze signals. In this context, we classify signal representations into two categories, namely, global and local representations, and this thesis revolves around these topics.

1.1 Global representations

A global representation of a time-signal is understood to mean a representation that involves the whole time-domain of the signal. This implies that in order to give a global representation of a signal, the signal has to be known completely (by measurements) and conversely, in order to evaluate the signal at a single moment in time from its global representation, in general, the representation has to be completely identified. A global representation of a signal is associated to a global spectrum, that is, the distribution of the global information of the signal towards the basic elements of the representation. Having a global representation of time-signals considered yields a one-to-one mapping of the description of the signals in the time-domain to their description in another domain, for instance, the frequency domain. In this sense, a global representation is complete.

As a simple example, we consider finite energy continuous-time signals in the Hilbert space $L_2(0, \infty)$, and take an orthonormal basis $\{\phi_n| n \in \mathbb{N}_0\}$ in $L_2(0, \infty)$. This orthonormal basis introduces a global representation. Indeed, every continuous-time signal $x$ can be represented as

$$x = \sum_{n \in \mathbb{N}_0} (\mathcal{Y}_i x)(n)\phi_n,$$  \hspace{1cm} (1.1)
where $\mathcal{Y}_1 : L_2(0, \infty) \longrightarrow \ell_2(\mathbb{N}_0)$ is the one-to-one mapping defined by

$$(\mathcal{Y}_1 x)(n) = \langle x, \phi_n \rangle, \ n \in \mathbb{N}_0. \quad (1.2)$$

The sequence $\{\langle x, \phi_n \rangle\}$ represents the global spectrum of $x$ associated to this type of representation. We note that to compute the spectrum $\mathcal{Y}_1 x$, complete knowledge of the signal $x$ is needed, and conversely, to represent $x$ from its spectrum $\mathcal{Y}_1 x$, this spectrum has to be known completely. The transformation $\mathcal{Y}_1$ in (1.2) is called a global transformation, and the representation in (1.1) is called a global representation of the signal $x$. The global spectrum corresponding to $x$ is the sequence $\langle x, \phi_n \rangle$, that is, the distribution of the information (energy) of $x$ over the basic elements $\phi_n, n \in \mathbb{N}_0$.

Another example is the discrete-time Fourier representation for discrete-time signals in $\ell_2(\mathbb{Z})$, that is,

$$x(t) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} (\mathcal{Y}_2 x)(e^{\omega t}) e^{\omega t} d\omega, \ t \in \mathbb{Z}, \quad (1.3)$$

where $\mathcal{Y}_2 : \ell_2(\mathbb{Z}) \longrightarrow L_2(-\pi, \pi)$ is the discrete-time Fourier transformation defined by

$$(\mathcal{Y}_2 x)(e^{\omega t}) = \frac{1}{\sqrt{2\pi}} \sum_{t \in \mathbb{Z}} x(t) e^{-\omega t}, \ \omega \in [-\pi, \pi]. \quad (1.4)$$

For this representation, $(\mathcal{Y}_2 x)(e^{\omega t}), \omega \in [-\pi, \pi]$, is the global spectrum of $x$. We observe that in literature $|\mathcal{Y}_2 x(e^{\omega t})|$ is called the global amplitude spectrum of $x$ and $\arg \{(\mathcal{Y}_2 x)(e^{\omega t})\}$ the global phase spectrum of $x$. To determine the global spectrum $(\mathcal{Y}_2 x)(e^{\omega t})$, the full information of the signal $x$ is needed and, also, to represent the signal for a single moment of time $t$ from its spectrum $(\mathcal{Y}_2 x)(e^{\omega t})$, this spectrum has to be known completely. The global transformation $\mathcal{Y}_2$ maps $\ell_2(\mathbb{Z})$ in $L_2(-\pi, \pi)$. So the signal $x$ is completely characterized if $(\mathcal{Y}_2 x)(e^{\omega t})$ is known for all $\omega \in [-\pi, \pi]$.

These two examples give an illustration of two types of global transformations. Both global transformations, $\mathcal{Y}_1$ and $\mathcal{Y}_2$, characterize the presence (and in which amount) of $\phi_n, n \in \mathbb{N}_0$, and of $e^{\omega t}, \omega \in [-\pi, \pi]$, respectively, in the signal $x$. The second global representation, though, has the additional property that it is linked to a physical characteristic of the signal, namely, its frequency domain. Transformation to another domain can be useful to understand or interpret a signal. Of course we can regard the transformation in (1.2) from $L_2(0, \infty)$ onto $\ell_2(\mathbb{N}_0)$ as a domain transformation, however, $\mathcal{Y}_1$ has no physical interpretation as long as the $\phi_n, n \in \mathbb{N}_0$, are not interpretable in this sense. Both types of global transformations can be used for analysis of signals.

“Analysis of a signal is the tracing of the signal to its source and the resolving of knowledge into its original principles.”

From dictionary.com
1.1 Global representations

Parameterized global representations

For clarity, we consider the representation in (1.1). The global spectrum $\mathcal{Y}_1x$ gives a distribution of the energy according to the orthonormal basis $\{\phi_n|n \in \mathbb{N}_0\}$ that is used. For practical purposes such as approximation, system identification and control, the representation is efficient if the distribution is condensed in only a small number of the elements of the basis. However, such a distribution varies from signal to signal. Therefore, in order to adapt to this problem, we consider instead of one basis $\{\phi_n|n \in \mathbb{N}_0\}$, a parameterized collection $\{\phi_n[\theta]|n \in \mathbb{N}_0\}$. As an example we consider the Laguerre functions:

$$\phi_n(t) = e^{-\frac{1}{2}t} \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{t^k}{k!}$$

for $n \in \mathbb{N}_0$, which establish an orthonormal basis in the Hilbert space $L_2(0, \infty)$. From the basis $\{\phi_n|n \in \mathbb{N}_0\}$, we obtain a parameterized collection $\{\phi_n[\theta]|n \in \mathbb{N}_0\}$ of orthonormal bases; with

$$\phi_n(t;\theta) = \sqrt{\theta} e^{-\frac{\theta x}{2}} \sum_{k=0}^{n} (-1)^k \binom{n}{k} \left(\frac{\theta x}{2}\right)^k \frac{t^k}{k!}, \quad n \in \mathbb{N}_0,$$

for $\theta > 0$. In fact, $\theta$ is a scaling parameter. Varying $\theta$ means stretching or shrinking. Given a finite energy signal $x$ in $L_2(0, \infty)$ we may optimize $\theta$ in such a way that the energy of $x$ is condensed in a small number of the elements of the basis. Let us consider the signal $x \in L_2(0, \infty)$:

$$x(t) = e^{-\frac{t}{2}}, \quad t \geq 0.$$

This signal can be represented in the basis $\{\phi_n|n \in \mathbb{N}_0\}$ as follows:

$$x(t) = \sum_{n=0}^{\infty} \langle x, \phi_n \rangle \phi_n(t), \quad t \geq 0,$$

where, by straightforward calculation, the coefficients are

$$\langle x, \phi_n \rangle = \frac{20}{11} \left( -\frac{9}{11} \right)^n, \quad n \in \mathbb{N}_0.$$

The relative energy loss by taking a truncation after $N$ terms is given by

$$\frac{\|x - \sum_{n=0}^{N-1} \langle x, \phi_n \rangle \phi_n\|^2}{\|x\|^2} = \left( -\frac{9}{11} \right)^{2N}.$$
So for a 95% accuracy, at least 15 terms are needed in the approximation. However, by making use of the parameterized collection \( \{ \phi_n[\theta] | n \in \mathbb{N}_0 \} \), one term is sufficient for the complete representation of \( x \),

\[
x(t) = \sqrt{10} \phi_0(t; \frac{1}{10}).
\]

In this thesis, we present a large class of orthonormal systems, which can be embedded into a parameterized collection of orthonormal bases, and given a signal \( x \) optimal parameters can be obtained yielding an orthonormal system such that the representation \( x \) in this system is condensed in a small number of the elements of the system.

### 1.2 Local representations

A local representation of a signal is understood to be a representation that characterizes properties of the signal on selected intervals which are bounded or effectively bounded. In contrast to global representations, local representations have the property that they focus on bounded time-intervals. To represent the restriction of a signal to a bounded time-interval, the distribution of the information of the signal towards the basic elements characterizing the representation is not completely needed. In fact, when we carry out an analysis with a local representation we always have a certain segmentation in mind. A local representation associates a signal to a local spectrum, that is, the local information distribution of the signal over the representation. The mapping that associates a signal with its local spectrum is called a local transformation. Local transformations can be used for local analysis, that is, an analysis directed to the following questions: What kind of features of the representation does the signal contain? How much of each feature is present in the signal? At which location is the feature present in the signal? In this thesis, we consider two types of local transformations, namely, local transformations by windowed unitary transformations and local transformations by orthogonal segmenting projections.

#### Local analysis by windowed unitary transformations

We explain this type of local analysis by means of an example. To this end, we consider the global transformation \( \mathcal{Y} \) defined by,

\[
(\mathcal{Y}x)(n) = \langle x, \phi_n \rangle, \quad n \in \mathbb{N}_0,
\]

where \( \{ \phi_n | n \in \mathbb{N}_0 \} \) is an orthonormal basis in the Hilbert space \( \ell_2(\mathbb{Z}) \). We introduce a real-valued window \( w \), that is, a real-valued sequence of the same type as \( x \), with a finite support (that is, \( w(t) \) is zero except for a finite number of \( t \)). In this thesis, we
consider also cases of \( w \) with infinite support but effectively bounded. We multiply a shifted version of the signal \( x \), by the window \( w \), and then we apply the global transformation to the windowed signal. In this way, we obtain a transformation \( \mathcal{W} \) defined by

\[
(\mathcal{W}x)(n, a) = \langle x(\cdot - a)w, \phi_n \rangle,
\]

for \( n \in \mathbb{N}_0 \). For determining \( (\mathcal{W}x)(n, a) \) for a fixed \( (n, a) \), we only need to know \( x \) on support\( \{ w(\cdot + a) \} \), which is finite. Since for all \( t \in \mathbb{Z}, a \in \mathbb{Z} \):

\[
x(t - a)w(t) = \sum_{n \in \mathbb{N}_0} (\mathcal{W}x)(n, a)\phi_n(t),
\]
or, equivalently,

\[
x(t)w(a) = \sum_{n \in \mathbb{N}_0} (\mathcal{W}x)(n, a - t)\phi_n(a),
\]

we have for all \( a \in \text{support}(w) \),

\[
x(t) = \frac{1}{w(a)} \sum_{n \in \mathbb{N}_0} (\mathcal{W}x)(n, a - t)\phi_n(a).
\] (1.5)

This means that, to represent \( x(t) \), for a single \( t \), out of \( \mathcal{W}x \), we do not need to know the complete spectrum \( (\mathcal{W}x)(n, a) \) for all \( a \in \mathbb{Z} \), but only for \( (n, a - t), n \in \mathbb{N}_0 \) and a single \( a \in \text{support}(w) \). The transformation \( \mathcal{W} \) is called a unitary windowed transformation, \( \mathcal{W}x \) is the local spectrum of \( x \) associated to \( \mathcal{W} \). The representation in (1.5) is a local representation of \( x \) associated to \( \mathcal{W} \). As we see from this example, a local transformation introduces redundancy and, therefore, possibilities to compress. So the question arises to characterize subsets \( \Gamma \) of \( \mathbb{N}_0 \times \mathbb{Z} \) such that having knowledge of \( \mathcal{W}x \) on \( \Gamma \) renders the possibility to represent \( x \).

A nice way of looking at this type of local analysis is by considering it as filtering; then we write

\[
(\mathcal{W}x)(n, a) = (A_n \ast x)(-a),
\]

where the \( A_n, n \in \mathbb{N}_0 \),

\[
A_n(t) = w(-t)\phi_n^*(-t),
\]

are called the analysis filters. The type of filtering as arising from the windowed unitary transformation can be seen as a system of parallel filters, which is called an analysis filter bank.
Local analysis by orthogonal segmenting projections

Given a finite energy signal \( x \in L^2(\mathbb{R}) \), and given a disjoint segmentation of the real line \( \{ I_k | k \in \mathbb{Z} \} \), that is, \( \cup_{k \in \mathbb{Z}} I_k = \mathbb{R} \) and \( I_k \cap I_l = \emptyset \) for \( k \neq l \). Let \( \chi_{I_k} \) define the characteristic function of the interval \( I_k \). Choose an orthonormal basis \( \{ \phi_{n,k} | n \in \mathbb{N}_0 \} \) in \( \chi_{I_k} L^2(\mathbb{R}) = L^2(I_k) \), \( k \in \mathbb{Z} \), then we obtain the local transformation \( \mathcal{W} \)

\[
(\mathcal{W}x)(n,k) = \langle x, \phi_{n,k} \rangle, \quad n \in \mathbb{N}_0, k \in \mathbb{Z}.
\]

The local representation of a signal \( x \) associated to the local transformation \( \mathcal{W} \) is given by

\[
x(t) = \sum_{k \in \mathbb{Z}} \sum_{n \in \mathbb{N}_0} \langle x, \phi_{n,k} \rangle \phi_{n,k}(t), \quad t \in \mathbb{R}.
\]

We note that

\[
\oplus_{k \in \mathbb{Z}} \chi_{I_k} L^2(\mathbb{R}) = L^2(\mathbb{R}),
\]

and that the collection \( \cup_{k \in \mathbb{Z}} \{ \phi_{n,k} | n \in \mathbb{N}_0 \} \) constitutes an orthonormal basis for \( L^2(\mathbb{R}) \). For speech signals, one often assumes stationarity on short time-intervals (of about 10ms). Then this segmentation can be of practical importance. However, the local representation based on this segmentation will suffer from boundary artifacts. So this concept has to be improved in order to suppress the boundary artifacts. The artifacts are caused by the discontinuity of the characteristic function at the boundaries. To avoid this discontinuity, instead of the characteristic function \( \chi_{I_k} \), we can use a function with a smooth cutoff at the boundaries of the interval \( I_k \) with a support overlapping with the intervals \( I_{k-1} \) and \( I_{k+1} \). Having such a smooth function \( \psi_k \), subject to some additional conditions, \( \psi_k \) induces an orthogonal projection \( \mathcal{P}_k \). Thus we obtain a local representation and a direct segmentation of \( L^2(\mathbb{R}) \) as in (1.7),

\[
\oplus_{k \in \mathbb{Z}} \mathcal{P}_k L^2(\mathbb{R}) = L^2(\mathbb{R}).
\]

Choosing orthonormal basis \( \{ \phi_{n,k} | n \in \mathbb{N}_0 \} \) for \( \mathcal{P}_k L^2(\mathbb{R}) \), \( k \in \mathbb{Z} \), the collection \( \{ \phi_{n,k} | n \in \mathbb{N}_0, k \in \mathbb{Z} \} \) constitutes an orthonormal basis for \( L^2(\mathbb{R}) \), and herewith a smooth orthogonal local representation is achieved.

1.3 Applications

System modeling and data fitting

In system identification one tries to represent the impulse response or transfer function of a system given measurements of inputs and corresponding outputs. In this
1.3 Applications

respect, there similarities to filter design where the objective is to represent filters
given requirements on the impulse response of the transfer function instead of mea-

An example can be found in the modeling of the auditory human system
using the Zwicker data [103]. The Zwicker data describes the relation between the
critical bands and the center frequencies of the auditory system. Using this data the
auditory system can be modeled by a Kautz representation, that is, a parameterized
global representation based on simple second-order systems, where the parameters
correspond to the critical bands and the center frequencies.

The first part of this thesis is devoted to the representation of signals in condensed
series expansions, with respect to parameterized orthonormal bases. The ideas de-
scribed in the first part of this thesis can be used, for example, as follows. Using
the Laguerre representation we measure the impulse response of a system accord-
ing to some requirements. Then we can calculate the optimal Laguerre parameter
to represent the impulse response in a condensed form, using these measurements.
This method leads to a condensed series representation, where only straightforward
numerical calculations on the impulse response of the system are required.

Digital communication

The idea of subdividing a signal frequency band into a set of contiguous bands of-
fers a powerful paradigm for accomplishing efficient digital communication across
band-limited communications channels. The principle was used more than 40 years
ago in the Collins Kineplex system, which transmitted data by subdividing it into
several interleaved bit streams, and using these to modulate several carriers [38].
In recent years, thanks in part to rapid advances in IC technology, various new ap-
lications have emerged in areas such as digital subscriber lines (e.g. ADSL and
VDSL [12, 26]) and wireless digital video and audio broadcast [40, 41, 71, 75], and
wireless local area networks [71]. Most of these applications use orthogonal fre-
quency division multiplexing (OFDM). This modulation technique is based on the
fast Fourier transform, can thus be implemented efficiently in silicon. FFT-based
filter banks, however, do not necessarily utilize the communication channel very ef-
ficiently, and a trend towards more powerful (and generally more complex) filter bank
modulation techniques can currently be discerned. This trend is supported by the
high cost of spectrum, and the ever decreasing cost of silicon. In Part II, we consider
analysis-synthesis systems leading to easily implementable representation schemes
by generalizing the concept of the short-time Fourier spectrum.

Audio representations

In digital audio processing, one is interested in audio signal representations that allow
for both high quality data compression at low bit rates and for compressed domain
processing. One way is to segment the audio signal in three signal components, namely, sinusoids, transients and noise. These three components are quantized separately and compressed before processing or modification. First one extracts the transients from the audio signal. For extracting sinusoids one uses multi-resolution modeling to encode the signal. In this multi-resolution modeling, two topics considered in this thesis may play a role, namely, the concept of filter bank and the concept of orthogonal segmenting projection. Having partitioned the audio signal into frames with respect to the time domain, the analysis filter bank decomposes the signal into a tuple of frequency band-limited signals. Each analysis filter is centered around a frequency band. So each of the frequency band-limited signals is linked to one or more center frequencies with corresponding amplitudes and phases, resulting in sinusoids. In this thesis we discuss how to segment sinusoids smoothly (without boundary artifacts), and give corresponding representations to the decompose the audio signal.

1.4 Goals of the thesis

We have divided this thesis into three main parts, namely, Part I on the analysis and design by global series representations, Part II on filter banks in a design perspective, and Part III on the analysis by domain segmentations. Now for each part we describe the underlying problems. The solutions which are proposed in this thesis, lead to the accomplishment of the goals we summarized below.

1.4.1 Analysis and design by global series representations

- To formulate mathematical criteria for different types of optimality of condensed representations of signals.

- To reformulate these mathematical criteria by making use of the fact that the basic elements satisfy a second-order ordinary differential/difference equation.

- To develop methods to solve the reformulated problem and, in particular, show how the methods work on the basis of examples.

- To consider the particular case of the Hermite representation which has nice properties in signal analysis.

- To give a conceptual solution for finding the optimal parameters of a Kautz representation, knowing that the system of Kautz functions does not correspond to a single second-order differential/difference equation.
1.5 Outline of the thesis

1.4.2 Filter banks in a design perspective

- To develop relations between windowed unitary transformations, filter banks and frames.
- Given a uniform subsampling scheme of the windowed unitary spectrum, to develop simple representation schemes.
- To introduce the concept of non-uniform subsampling and develop the corresponding representation schemes.
- To give illustrations on the basis of the Kautz and the Laguerre systems.

1.4.3 Analysis by domain segmentations

- Given a domain segmentation, to develop methods to suppress the artifacts at the transition boundaries, in particular for the one- and the two-dimensional domains.
- Given a signal, to determine the optimal segmentation of its underlying domain.
- Given a domain segmentation, to determine suitable collections of basic elements.

1.5 Outline of the thesis

This thesis consists of three parts:

P I. Analysis and design by global series representations.

P II. Filter banks in a design perspective.

P III. Analysis by domain segmentations.

In the first part we investigate the parameterized global representations. Chapter 2 gives an introduction to Part I. Firstly, we present a short historical overview on some specific well-known global representations. Secondly in this introduction, we present the idea of condensed representations by parameterized orthonormal bases, and thirdly we describe in detail the content of the first part. In Chapter 3, we define the problem of finding optimal parameters for the representation of a signal, given a parameterized global representation. This choice is based on an optimality criterion. We reformulate this criterion by making use of the fact that the orthonormal bases in the parameterized collection satisfy a certain second-order differential/difference
equation. In Chapter 4, we deal with the parameterized collection of the modulated Hermite bases. The optimal parameter obtained can be linked, to physically interpretable properties of the signal. Causal time-invariant linear systems can be approximated by finite Laguerre or Kautz systems. In Chapter 5, we treat this approximation problem along the method described in Chapter 3 regarding as the parameter a sequence of poles in the Laguerre and Kautz systems. We also show equivalence between discrete-time causal systems and continuous-time causal systems in obtaining the optimal parameters.

In the second part, we deal with local analysis by windowed unitary transformations for discrete-time signals. Chapter 6 gives an introduction to Part II. In this introduction, we discuss the link between a subsampled windowed unitary spectrum and a filter bank, we classify the subsampling schemes as being uniform or non-uniform, and we describe, in somewhat more detail, the content of the second part. In Chapter 7, we provide the uniformly subsampled case. The basic building blocks in a filter bank are presented and their properties are given. We deal with different concepts that play a role in a multirate system, such as causality and stability and perfect reconstruction. Further, we give representation schemes and we consider illustrations based on Laguerre and Kautz filters. The setup of Chapter 8 is similar to the one of Chapter 7, replacing uniform downsampling by a non-uniform one. So we define the representation problem, we reformulate it to a subsampled windowed unitary transform and to an ‘over’ complete system and give conditions for representation. A uniform downsampled analysis filter bank equivalent to the non-uniform one is created and a representation is given.

In the third part we deal with techniques for obtaining local representations by orthogonal segmenting projections. Chapter 9 gives an introduction to Part III. On the basis of examples, we show that straightforward segmentation by the characteristic functions causes boundary artifacts. How to overcome these artifacts is one of the topics of Chapter 10. For one-dimensional segmentation we deal with local representations based on segmentations for both discrete-time energy signals as well as for continuous-time energy signals. In Chapter 11, we consider the artifact reduction in the two-dimensional discrete energy signals. We discuss a straightforward generalization from the one-dimensional segmentation leading to a segmentation by rectangles. We show that domain segmentation by polygons without introducing artifacts is not possible in general. Only in the very special case of domain segmentation by a particular type of hexagons we are able to show that there exist corresponding orthogonal segmenting projections that do not introduce boundary artifacts.
1.6 Main results of the thesis

Part I:

- A mathematical formulation of selecting the optimal parameters for a parameterized class of orthogonal expansions more rigorous than in literature [20, 87].
- Application to Jacobi expansions and harmonically modulated Laguerre / Hermite expansions.
- A recursive procedure for selecting optimal poles in Kautz expansions.

Part II:

- Introduction of the concept of sliding window unitary transformation, generalizing the concept of sliding window Fourier transformation.
- Analysis of this type of transformation in the setting of filter banks and frames.
- Application to Laguerre and Kautz systems.

Part III:

- Analysis of the discrete-time lapped orthogonal segmentation.
- Description of a design methodology for (tight) frames on the basis of the lapped orthogonal segmentation.
- Description of a procedure to construct a lapped orthogonal hexagonal segmentation.
Part I

Analysis and design by global series representations
Chapter 2

Introduction to Part I

Identification and approximation of dynamical systems by a set of orthonormal functions are techniques often applied. The areas of importance include circuit and system theory, control system design [6, 23, 50, 85, 91, 94–96] and signal processing [21, 24, 29, 37, 67]. The impulse response of a physical system can often be approximated by exponential polynomials. In this context we mention that one of the famous techniques in orthogonal approximation is the Laguerre\(^1\) series expansion, which dates back to Wiener [100] and Lee [57]. Wiener observed that the classical Laguerre functions, in their Laplace domain, are useful for approximating linear dynamical systems and have a very nice recursive structure, see Fig. 2.1. These functions were introduced in 1879 by Laguerre. Lee implemented linear dynamical systems described by the Laguerre functions with an efficient RLC electrical network.

\[
\delta \rightarrow B(s) \rightarrow A(s) \rightarrow A(s) \rightarrow \cdots \\
\Phi_0[\sigma] \quad \Phi_1[\sigma] \quad \Phi_2[\sigma]
\]

*Figure 2.1: Recursive structure of the Laguerre system, where \(A(s)\) and \(B(s)\) are first-order systems.*

Later, in the fifties and the beginning of the sixties, there was further progress with this approximation theory. The problem of orthogonalizing a set of continuous-time exponential polynomials was elegantly solved by Kautz [55], and already earlier by Walsh [97], Takenaka [86] and Malmquist [65]. The Laplace transforms of the Kautz functions have also a very nice recursive structure providing simple implementation,

\(^1\)Edmond Laguerre lived from 1834 to 1886: attended the Ecole Polytechnique in Paris but only ranked 46th in his class. He was an artillery officer from 1854 to 1864 when he returned to the Ecole Polytechnique where he remained for the rest of his life.

Laguerre studied approximation methods and is best remembered for the special functions: the Laguerre polynomials.
see Fig. 2.2. Kautz functions are mainly applied in filter design and approximation of transient signals. A complete chapter in [52] is devoted to the description of using Kautz functions in approximation methods in feedback system design.

Recently there has been considerable interest in using Laguerre and Kautz functions in system theory, approximation and identification for control [8, 17, 23, 42, 44, 50, 63, 91]. For discrete-time dynamical systems, discrete versions of Laguerre and Kautz can be used [25].

Another famous technique in orthogonal approximations is the Hermite\textsuperscript{2} series expansions. The Hermite functions are eigenfunctions of the Fourier transformation and therefore play a special role in signal analysis [53]. In [36, 67] the Hermite series expansions of two variables are used for image analysis, where the properties of the overlapping Gaussian windows are used to project images locally onto the basis of Hermite polynomials. Since the analysis filters in the Hermite transformation are derivatives of Gaussians, see [68], the image analysis is in close agreement with the information analysis carried out by the human visual system. Other applications in pattern recognition and signal coding can be found in [21, 29, 37]. In [24] it is shown that the first-order moment of the energy in the Hermite domain of a signal is related to the time-bandwidth product. This way, a link to the work of Gabor is made. Therefore, we will compare the Gabor expansion with a modulated Hermite series expansion, see Chapter 4.

\textsuperscript{2}Charles Hermite lived from 1822 to 1901: always found formal examinations difficult and had to spend 5 years working for his B.Sc. which he received in 1848. He held posts at the École Polytechnique, Collège de France, École Normale Supérieure and the Sorbonne.

Hermite is known also for a number of mathematical entities that bear his name, Hermite polynomials, Hermite’s differential equation, Hermite’s formula of interpolation and Hermitian matrices.
2.1 The scope of Part I

Let the collection \( \{ \phi_n[\theta] \mid n \in \mathbb{N}_0 \} \), \( \theta \in \Theta \), be an orthonormal basis in a Hilbert space \( \mathcal{H} \), where \( \theta \in \Theta \) is a parameter vector and \( \Theta \) is the parameter set. The parameter \( \theta \) can be seen as a freedom introduced in a certain basis, for example, as a shift, a scale or a modulation parameter. Thus for each choice of the parameter \( \theta \) and each element \( x \) in \( \mathcal{H} \) we can write:

\[
x = \sum_{n=0}^{\infty} \langle x, \phi_n[\theta] \rangle \phi_n[\theta].
\]

In practice, the series will be truncated after a finite number of terms, say \( N \). Then an approximation \( x_N[\theta] \) of \( x \) emerges:

\[
x_N[\theta] = \sum_{n=0}^{N-1} \langle x, \phi_n[\theta] \rangle \phi_n[\theta].
\]

This approximation \( x_N[\theta] \) is the orthogonal projection of the function \( x \) onto the subspace \( M_N[\theta] = \text{span}\{\phi_n[\theta] \mid n = 0, \cdots, N-1\} \). Observe that \( M_N[\theta] \) can be defined as the subspace generated by any independent collection \( \{\psi_n[\theta] \mid n = 0, \cdots, N-1\} \) where \( \{\phi_n[\theta] \mid n = 0, \cdots, N-1\} \) results from the Gram-Schmidt orthonormalization procedure applied to \( \{\psi_n[\theta] \mid n = 0, \cdots, N-1\} \). The approximation \( x_N[\theta] \) depends on the free parameter \( \theta \). In Part I we investigate the following problem:

**Problem:** Consider a fixed function \( x \) in \( \mathcal{H} \), which is in general not completely known except for some experimental data. Determine \( \theta \) and \( N \) explicitly, without numerical search procedures, such that the distance between \( M_N[\theta] \) and \( x \) is sufficiently small.

Minimizing the quadratic error \( \mathcal{E}_{\text{Err}}(N, x; \theta) := \|x - x_N[\theta]\|^2 \) could be a natural problem definition. However, this problem is very difficult to solve. Following such an approach, costly search procedures have to be used [66] and the problem of the occurrence of local minima has to be solved. In practice, we would like to establish the parameters before actually engaging the orthonormal expansion, see Fig. 2.3. Therefore, we introduce an **enforced convergence criterion**, that is, an upper bound of the quadratic error that makes use of properties of the basic elements of the series expansion.

2.2 Optimal parameters

In Chapter 3, we consider an approach for solving the problem above, that can be useful in practice: the so-called enforced convergence criterion. We describe this ap-
proach and illustrate it by applying it to specific functions. Firstly, we introduce the class of functions that we will represent in the series expansions. Secondly, we derive an upper bound for the quadratic truncation error and introduce the enforced convergence criterion. Thirdly, we consider the optimization problem for the enforced criterion. It is based on the minimization of the quadratic form which depends on the parameter $\theta$. This quadratic form is fixed by a differential/difference operator $A[\theta]$, mainly of second-order, which is related to the orthonormal basis $\{\phi_n[\theta]|n \in \mathbb{N}_0\}$ such that $A[\theta]\phi_n[\theta] = \eta_n[\theta]\phi_n[\theta]$, for a non-negative sequence $\eta_n[\theta]$. This way, the enforced convergence criterion leads to simple expressions for the optimal parameters. The optimal solution does not require complete knowledge of the function considered, but only some experimental measurements. This gives us opportunities for design purposes. Furthermore, the upper bound may become tighter if we replace $A[\theta]$ by $A^k[\theta]$, $k \geq 2$.

In contrast to the quadratic error criterion, the enforced convergence criterion has the convenient property that it defines the optimal parameter, independent of the number of terms in the expansion. Illustrations for the enforced convergence criterion are given by computing the optimal parameters for the Jacobi\textsuperscript{1}, Charlier and Meixner expansions. In these cases, we give the explicit expressions for the optimal parameters as functions of measurements.

An interesting question is the following: given a certain choice of parameters, determine the class of functions that has these parameters as optimal solution. We

---

\textsuperscript{1}Carl Gustav Jacob Jacobi lived from 1804 to 1851: Jacobi’s early education was given by an uncle on his mother’s side, and then, just before his twelfth birthday, Jacobi entered the Gymnasium in Potsdam. He had been well taught by his uncle and he had remarkable talents, so in 1817, while still in his first year of schooling, he was put into the final year class. This meant that by the end of the academic year 1816-17 he was still only 12 years old yet he had reached the necessary standard to enter university. The University of Berlin, however, did not accept students below the age of 16, so Jacobi had to remain in the same class at the Gymnasium in Potsdam until the spring of 1821.

Jacobi carried out important research in partial differential equations of the first order and applied them to the differential equations of dynamics.
can treat this problem by putting the expressions for the optimal parameters equal to the values chosen. This way we can divide the admissible space into a collection of classes according to the values of the parameters. We give an illustration for the Chebyshev\(^4\) expansions, where the parameter is fixed.

2.3 Analysis by Hermite transformations

In Chapter 3 we could have used the Hermite instead of Jacobi expansions. For signal analysis we will see that by applying the enforced convergence criterion, we obtain a natural description for the optimal parameters. Hermite series expansions are often used as an analysis transformation for coding as well as for processing. To make coding and processing computationally efficient, we want that the analysis transformation is as compact as possible. Since many of these series are determined up to one or more degrees of freedom, a transformation can be compacted by choosing the parameters appropriately. The question is then how to do this. For this, the enforced convergence criterion as a compaction criterion is defined. Obviously, having a compaction criterion, we could do a large-scale search over the complete degrees of freedom and, thus, establish the optimum compaction parameters according to the given criterion. This is of course computationally rather expensive. It would be much more convenient to have a closed-form expression for the optimum compaction parameters, avoiding any search procedure.

The harmonically modulated Hermite functions constitute an orthonormal basis in the Hilbert space of square-integrable functions \(L^2(\mathbb{R})\). This basis comprises three free parameters, namely a translation, a modulation, and a scale factor.

We will introduce a compaction criterion [20, 87], as addressed in the previous chapter for the enforced convergence criterion, and show that this specific measure results in simple, explicit expressions for the free parameters in a modulated Hermite series expansion as a function of the signal which is to be expanded. Moreover, the results have a clear physical interpretation, specifically, the optimum translation equals the first-order moment of the energy distribution in the time domain, the optimum modulation equals the first-order moment in the frequency domain, and the squared optimum scale equals the geometrical mean of the second-order central moment in the frequency domain and the inverse of the second-order central moment in the time domain. This result extends earlier work in this area; see [20] and references therein. This chapter has been published in [22, 24].

---

\(^4\) Pafnuty Lvovich Chebyshev lived from 1821 to 1894: In 1847 Pafnuty Chebyshev was appointed to the University of St Petersburg. He became a foreign associate of the Institut de France in 1874 and also of the Royal Society.

He wrote about many subjects, including probability theory, quadratic forms, orthogonal functions, the theory of integrals, the construction of maps, and the calculation of geometric volumes.
20 Introduction to Part I

2.4 Design of causal systems

Thinking of the use of Laguerre and Kautz functions, we end up naturally in problems of system identification [94, 95] and signal coding [21]. For system identification, we are mainly interested in orthogonal series expansions on a semi-infinite axis. Restricting our attention to rational transfer functions, we have orthogonalized exponential polynomials [98] as the basis blocks. Applying the Gram-Schmidt orthonormalization procedure to the exponential polynomials, we obtain the so-called Kautz functions [25, 55], with as a specific case the well-known Laguerre functions [57].

In Chapter 5, we introduce the modulated Laguerre expansions. In the Laplace domain they are related to Laguerre functions that allow complex poles. The modulated Laguerre expansions can be used in approximating functions with a clearly resonant character. In the first section of this chapter, we derive properties of the expansions and compute the optimal parameters according to the enforced convergence criterion.

Having the optimal pole $p = -\frac{1}{2} \sigma + j \Omega$ for the modulated Laguerre expansions, see Fig. 2.1, the Laguerre series describes compactly a function that is to be approximated, if it has a single side-band modulated low-pass nature. However, on qualitative grounds one can expect that Kautz series of more general type achieve an even more compact representation in the case of approximating a bandpass filter.

In second section of this chapter, we consider the Kautz expansions. The modulated Laguerre system is special case of a Kautz system. First, we show the similarity between the continuous-time and discrete-time. In this section, we concentrate mainly on discrete-time. A general Kautz system does not adhere to a fixed difference/differential equation. So the optimization problem that occurs has to be treated in a different way. In this section, we propose two methods for optimizing the parameters.

The first method is useful for functions with clearly resonant character. Therefore we restrict to a Kautz series having a repeated complex-conjugated pole pair. From the function $x$ we construct the function $y$ by the Hilbert transformation $x = (y + y^*)/2$. Next, we use this complex function $y$ to find the optimal parameter in a complex Laguerre series. Subsequently, we take this pole, and its complex conjugate, as an estimate for the optimal parameters in the Kautz series.

The second method is the recursive enforced convergence criterion, see Fig. 2.4. It determines the suboptimal parameters recursively: given a function $x$. We determine first the optimal parameter $\hat{p}$ for $x$ in the Laguerre system. Next, we consider the Kautz system where $p_k = \rho$, for $k \geq 1$ and $p_0 = \hat{p}$. This system adheres a difference/differential equation, where the optimal $\rho$ can be determined according to the enforced convergence criterion. Subsequently, we put $p_0 = \hat{p}$, $p_1 = \hat{p}$, $p_k = \rho$, $k \geq 1$, and so on we determine the next $p_k$. Recursively, we have a suboptimal solution for the Kautz system. Of course, combinations of the first and second method
can be taken.

As illustrations we give some academic examples. A part of this chapter has been published in [23, 81].

“When I consider what people generally want in calculating, I found that it always is a number. I also observed that every number is composed of units, and that any number may be divided into units. Moreover, I found that every number which may be expressed from one to ten, surpasses the preceding by one unit: afterwards the ten is doubled or tripled just as before the units were: thus arise twenty, thirty, etc. until a hundred: then the hundred is doubled and tripled in the same manner as the units and the tens, up to a thousand; ... so forth to the utmost limit of numeration.”

Abu Ja‘far Muhammad ibn Musa Al-Khwarizmi
Chapter 3

Optimal parameters in orthonormal series expansions

In the last decade with an increasing interest, orthogonal expansions have been used in approximation, coding, filtering, signal analysis, system modeling, system identification and system control [6, 21, 29, 37, 67, 85]. For these purposes as well as for computationally efficient processing, it is desired that the representation is as compact as possible. Obviously, having a compaction criterion, we could do a large-scale search over the complete degrees of freedom and, thus, establish the optimum compaction parameters according to the given criterion. This is of course computationally rather expensive. In general, in the quadratic error criterion to be optimized in order to compact the representation, the free parameters appear in a nonlinear, complicated form. This means that they are hard to find: different local solutions may occur from an optimization procedure and in fact, refined costly search procedures have to be used [66]. It would be more convenient to have a closed-form expression for the optimum compaction parameters, avoiding any search procedure, before actually engaging the orthogonal expansions. The question is then how this can be achieved. An approach to solve this question was presented in [27, 42, 48, 74] where the problem was restricted to Laguerre expansions. In [87] this approach was generalized to real-valued orthonormal functions adhering to a linear second-order differential equation for continuous-time or difference equation for discrete-time. In a recent work [20] this approach has been extended to higher-order differential or difference equation.

The specific form of the differential or difference equation results in an upper bound for the quadratic truncation error as well as an ‘interpretable’ expression for an enforced convergence rate criterion. These expressions depend on the free parameters and signal measurements. The enforced criterion can be optimized and explicit formulas for the optimal parameters of the classical orthogonal series expansions evolve.

In this chapter, we consider the enforced convergence criterion approach; we describe this approach and illustrate it by applying it to specific functions. Firstly, we introduce the class of functions that we will represent in the series expansions.
Secondly, we derive the upper bound for the quadratic truncation error and introduce the enforced convergence criterion. Thirdly, we consider the optimization problem for the enforced criterion. Further, we consider differential and difference operators and derive properties for the enforced criterion using these operators. In Sections 3.3 and 3.4 we illustrate the method. In Section 3.3, we consider the Jacobi functions for the continuous-time case. In Section 3.4, we consider the Charlier and Meixner functions for the discrete-time case. Finally, in Section 3.5, we conclude the chapter with a discussion.

3.1 Optimality condition

Typical sets of functions considered in this chapter are signals, system impulse responses or transfer functions. These sets are embedded in Hilbert spaces. Therefore we use orthogonal expansions to represent the functions in these sets.

The classical orthogonal expansions contain free parameters, for example, the pole in the Laguerre functions is the free parameter in Laguerre series expansions. For a start, let the collection \( \{\phi_n[\theta]|n \in \mathbb{N}_0 \} \), \( \theta \in \Theta \), be an orthonormal basis in a Hilbert space \( \mathcal{H} \), where \( \theta \in \Theta \) is a parameter vector (of one or more entries) and \( \Theta \) is the parameter set, that is, the collection of all parameter vectors \( \theta \) under consideration. In general, \( \Theta \) will be a subset of \( \mathbb{C}^p \), \( p \in \mathbb{N} \cup \{\infty\} \). Thus for each choice of the parameter \( \theta \) and each element \( x \) in \( \mathcal{H} \) we can write:

\[
x = \sum_{n=0}^{\infty} \langle x, \phi_n[\theta] \rangle \phi_n[\theta].
\]

If used in practice, the series will be truncated after a finite number of terms, say \( N \). Then an approximation \( x_N[\theta] \) of \( x \) evolves:

\[
x_N[\theta] = \sum_{n=0}^{N-1} \langle x, \phi_n[\theta] \rangle \phi_n[\theta].
\]

This approximation \( x_N[\theta] \) is the orthogonal projection of the function \( x \) onto the subspace \( \mathcal{M}_N[\theta] = \text{span}\{\phi_n[\theta]|n = 0, \ldots, N-1\} \). The approximation \( x_N[\theta] \) depends on the free parameter \( \theta \). The approximation \( x_N[\theta] \) is adequate if the distance \( \text{dist}(x, \mathcal{M}_N[\theta]) := ||x - x_N[\theta]|| \) between \( \mathcal{M}_N[\theta] \) and \( x \) is sufficiently small, for fixed \( x \) and \( N \). From the orthogonality of \( \{\phi_n[\theta]|n \in \mathbb{N}_0\} \) and the Pythagoras theorem we know that:

\[
||x - x_N[\theta]||^2 = ||x||^2 - ||x_N[\theta]||^2 = \sum_{n=N}^{\infty} |\langle x, \phi_n[\theta] \rangle|^2.
\]
3.1 Optimality condition

We denote the quadratic error by the expression $\mathcal{E}_{rr}(N, x; \theta) := \|x - x_N[\theta]\|^2$. Physically, this error may represent the energy of $x$ in the remaining part of the series expansion, that is, $\sum_{n=N}^{\infty} \langle x, \phi_n[\theta] \rangle \phi_n[\theta]$. The natural question is whether we can find an optimal parameter $\theta$ minimizing the error $\mathcal{E}_{rr}[\theta]$ (for fixed $x$ and $N$), with respect to this parameter $\theta$. The error $\mathcal{E}_{rr}[\theta]$ is then the criterion to be optimized. The parameter $\theta(x, N)$ obtained by the minimization of $\mathcal{E}_{rr}[\theta]$ is called the optimal parameter (later on we give another definition for the optimality) and $x_N[\theta]$ is the optimal truncation of $x$ by $N$ terms.

Minimization of the criterion $\mathcal{E}_{rr}(N, x; \theta) = \sum_{n=N}^{\infty} |\langle x, \phi_n[\theta] \rangle|^2$ with respect to $\theta$ is a very difficult problem to solve: if we are able to compute $\sum_{n=N}^{\infty} |\langle x, \phi_n[\theta] \rangle|^2$, then several local minima may occur. Following such an approach, costly search procedures have to be used to solve this problem (see [66] and references therein). To avoid such search procedures, we need alternatives for establishing an optimal $\theta$.

For classical orthogonal polynomials [62], we show that an alternative optimization problem can be defined, with the criterion,

$$
Q(x; \theta) := \sum_{n=0}^{\infty} \eta_n[\theta]|\langle x, \phi_n[\theta] \rangle|^2,
$$

where $(\eta_n)_{n \in \mathbb{N}_0}$ is a non-negative sequence, with the property $1/\eta_n[\theta] = O(1/n^p)$, $(n \to +\infty), p \geq 1$. In this case not all elements of $\mathcal{H}$ are taken into account, but a subspace $\mathcal{D} \subseteq \mathcal{H}$. The subspace $\mathcal{D}$ is called the admissible space for $\{\phi_n[\theta]|n \in \mathbb{N}_0\}$:

**Definition 3.1.1** Let $(\eta_n)_{n \in \mathbb{N}_0}$ be a non-negative sequence with the asymptotical property

$$
\frac{1}{\eta_n[\theta]} = O\left(\frac{1}{n^p}\right), \quad (n \to +\infty),
$$

where $p \geq 1$. Further, let $\{\phi_n[\theta]|n \in \mathbb{N}_0\}$ be an orthonormal basis in a Hilbert space $\mathcal{H}$. The admissible space $\mathcal{D}$ for $\{\phi_n[\theta]|n \in \mathbb{N}_0\}$ is defined by:

$$
x \in \mathcal{D} \iff x \in \mathcal{H} \land \forall \theta \in \Theta \sum_{n=0}^{+\infty} \eta_n[\theta]|\langle x, \phi_n[\theta] \rangle|^2 < +\infty.
$$

(3.1)

It can be shown that the admissible space $\mathcal{D}$ is dense in $\mathcal{H}$, and that all $\phi_n[\theta] \in \mathcal{D}$. In continuous-time cases, we will use the admissible space by $\mathcal{D}_c$ and in discrete-time cases by $\mathcal{D}_d$. Further, the set $\mathcal{D}^*$ denotes $\mathcal{D}$ excluding the zero-function.
Lemma 3.1.2 There exist $N_0 \in \mathbb{N}$ and $C > 0$ such that:

$$\forall x \in \mathcal{D} \forall N \geq N_0 \quad \|x - x_N[\theta]\|^2 \leq \frac{1}{CN^p} \sum_{n=0}^{\infty} \eta_n[\theta]\langle x, \phi_n[\theta]\rangle^2. \quad (3.2)$$

Proof: There exist $N_0 \in \mathbb{N}$ and $C > 0$ such that for all $N \geq N_0$

$$\eta_n[\theta] \geq CN^p, \text{ for } n \geq N.$$ 

Thus for $N \geq N_0$:

$$\sum_{n=0}^{\infty} \eta_n[\theta]\langle x, \phi_n[\theta]\rangle^2 \geq \sum_{n=N}^{\infty} \eta_n[\theta]\langle x, \phi_n[\theta]\rangle^2 \geq CN^p \sum_{n=N}^{\infty} \|\langle x, \phi_n[\theta]\rangle\|^2 = CN^p \|x - x_N[\theta]\|^2.$$ 

The expression $Q(x; \theta) = \sum_{n=0}^{\infty} \eta_n[\theta]\langle x, \phi_n[\theta]\rangle^2$ constitutes an upper bound for the quadratic error:

$$\mathcal{E}(N, x; \theta) \leq CQ(x; \theta)/N^p. \quad (3.3)$$

Of course, this makes sense if $CQ(x; \theta)/N^p \leq \|x\|^2$. The expression $Q[\theta]$ is independent of $N$. If $\eta_n = n^p$, it represents the $p$-th order moment of the energy of $x$ in the transfer domain.

To make this upper bound $CQ(x; \theta)/N^p$ as tight as possible, we seek the value of $\theta$ minimizing $Q[\theta]$. Instead of minimizing the quadratic error $\mathcal{E}(N, x; \theta)$, we minimize the upper bound $Q(x; \theta)$. Choosing the value of the parameter $\theta$ minimizing $Q(x; \theta)$, we have an indication of the required number of terms $N$ prior to the execution of the expansion. We make use of the inequality (3.3) in the following way. We compute the optimal parameter by minimizing of $Q(x; \theta)$ with respect to $\theta$. Then we see that we have to choose:

$$N \geq \sqrt[2p]{CQ(x; \theta)/\epsilon}, \quad (3.4)$$

in order to obtain a quadratic error $\mathcal{E}(N, x; \theta) \leq \epsilon$.

As known from literature [62], most of the classical orthogonal functions adhere to a finite-order ordinary differential or difference equation and, accordingly, we
3.1 Optimality condition

assume that the functions \( \phi_n[\theta] \) are eigenfunctions of an operator \( \mathcal{A}[\theta] \) with corresponding eigenvalues \( \eta_n[\theta] \):

\[
\mathcal{A}[\theta] \phi_n[\theta] := \eta_n[\theta] \phi_n[\theta], \quad n \in \mathbb{N}_0.
\] (3.5)

The operator \( \mathcal{A}[\theta] \) is defined on the admissible space \( \mathcal{D} \). Given this operator, the expression of \( Q(x; \theta) \) can be simplified to:

\[
Q(x; \theta) = \sum_{n=0}^{\infty} \langle x, \phi_n[\theta] \rangle \langle \mathcal{A}[\theta] \phi_n[\theta], x \rangle
= \langle \mathcal{A}[\theta] x, x \rangle.
\] (3.6)

As we will see next, in this way, \( Q(x; \theta) \) becomes a simple expression with a finite number of terms, that is, a sum of operators operating on \( x \).

If the parameter \( \theta \) can be separated from the \( x \) (from the inner product in (3.6)), the expression becomes simpler and so becomes the minimization of \( Q(x; \theta) \), for a fixed \( x \in \mathcal{D} \). This is the starting point of the next section.

For eventually a tighter upper bound for the quadratic error we suggest the following. The functions in the system \( \{ \phi_n[\theta] | n \in \mathbb{N}_0 \} \), (3.5), are also eigenfunctions of the differential/difference operator \( \mathcal{A}[\theta] := (\mathcal{A}[\theta])^q \), but now with eigenvalues \( (\eta_n[\theta])^q \).

So we can use as a criterion \( Q_q(x; \theta) = \sum_{n=0}^{\infty} (\eta_n[\theta])^q |\langle x, \phi_n[\theta] \rangle|^2 \), instead of (3.6), where the upper bound (3.4) for the new enforced convergence criterion is:

\[
N \geq \frac{\sqrt{C Q_q(x; \theta)}}{\varepsilon},
\]

This can be used to obtain a tighter upper bound for the quadratic error. However, for higher values of \( q \), this may lead to a more complicated expression of the enforced convergence criterion \( Q_q(x; \theta) \).

3.1.1 Continuous-time case

Let the system \( \{ \phi_n[\theta] | n \in \mathbb{N}_0 \} \) be an orthonormal basis in the Hilbert space \( \mathcal{H} \), which is a space consisting of functions of a continuous-time variable \( t \). The general form of the linear differential operator \( \mathcal{A}[\theta] \) is given by:

\[
\mathcal{A}[\theta] = \sum_{k=0}^{K} A_k(t; \theta) \left( \frac{d}{dt} \right)^k,
\] (3.7)

where \( K > 0 \). The function \( \phi_n[\theta] \) are eigenfunctions of \( \mathcal{A}[\theta] \) with eigenvalues \( \eta_n[\theta] \), for \( n \in \mathbb{N}_0 \). Examples of differential operators \( \mathcal{A}[\theta] \) for classical orthogonal functions
are given in subsequent sections. Substituting (3.7), the expression for \( A[\theta] \), into (3.6) leads to:

\[
Q(x; \theta) = \sum_{k=0}^{K} \langle A_k(t; \theta) \left( \frac{d}{dt} \right)^k x, x \rangle.
\] (3.8)

For the classical orthogonal functions, we shall prove in the next Section that the functions \( A_k(t; \theta) \) of \( t \) and \( \theta \) can be decomposed as a sum of products of two functions, of \( t \) and \( \theta \), separately, that is, each coefficient in the differential operator can be written as:

\[
A_k(t; \theta) = \sum_{j=1}^{J_k} B(k, \theta)C_{k, j}(t),
\] (3.9)

for some \( J_k \in \mathbb{N} \) and for \( k = 0, \ldots, K \). By an example, we will see in the next section that the representation in (3.9) is not unique. In this case, we choose a combination of \( B(k, \theta) \)’s and \( C_{k, j} \)’s that gives a condensed representation of \( A_k \). Substituting the expression (3.9) of \( A_k \) into (3.8), we obtain:

\[
Q(x; \theta) = \sum_{k=0}^{K} \sum_{j=1}^{J_k} B(k, \theta)\langle C_{k, j}(t) \left( \frac{d}{dt} \right)^k x, x \rangle,
\]

\[
= \sum_{k=0}^{K} \sum_{j=1}^{J_k} B(k, \theta)\mathcal{M}_{k, j}(x).
\]

Thus we have in \( Q(x; \theta) \) a separated form of the parameter \( \theta \) and certain functionals of \( x \). The expressions \( \mathcal{M}_{k, j}(x) := \langle C_{k, j}(d/dt)^k x, x \rangle \) are measurements on the function \( x \) and are independent of the parameter \( \theta \), which means that the \( \mathcal{M}_{k, j}(x) \)’s are completely determined if \( x \) is known. In the optimization scheme to be worked out further, the values \( \mathcal{M}_{k, j}(x) \) are taken into account as the only information retrieved from \( x \). So all \( x \) with the same \( \mathcal{M}_{k, j}(x) \), for all \( j \) and \( k \), yield the same optimal choice of \( \theta \). The admissible space \( D \) introduced in (3.1) can be replaced by \( D_c \), defined by:

\[
D_c = \{ x \in \mathcal{H} | |\mathcal{M}_{k, j}(x)| < +\infty, k = 0, \ldots, K; j = 1, \ldots, J_k \}.
\] (3.10)

We note that differentiability of \( x \in D_c \) is a condition implicitly contained in the definition of \( D_c \).

**Minimization Problem 3.1.3** For \( x \in D_c, x \neq 0 \), find \( \theta \in \Theta \) minimizing:

\[
Q(x; \theta) = \sum_{k=0}^{K} \sum_{j=1}^{J_k} B(k, \theta)\mathcal{M}_{k, j}(x),
\]
3.1 Optimality condition

where $\theta$ is the compaction parameter, $\mathcal{M}_{k,j}(x)$ is a measurement on $x$ and $Q[\theta]$ is the enforced convergence criterion to be optimized. The criterion $Q[\theta]$ is also called the compaction criterion. The optimal value of $\theta$ is denoted by $\hat{\theta}$.

Assuming that the functions $B_{k,j}$ are differentially dependent on the parameter $\theta$, the candidates for the optimality are obtained from the boundary of $\Theta$ and from:

$$\nabla_{\theta} Q[\theta] = \sum_{k=0}^{K} \sum_{j=1}^{J_k} \mathcal{M}_{k,j}(x) \nabla_{\theta} B_{k,j}(\theta) = 0,$$

(3.11)

for a fixed $x$ in $D_c$. Thus, optimization of $Q[\theta]$ results in an expression of the optimal $\theta$ in terms of the quantities $\mathcal{M}_{k,j}(x)$, see the next sections.

3.1.2 Discrete-time case

The discrete-time case is quite similar to the continuous-time case. The time is discrete, but the parameters are continuous. The difference is that we deal here with a difference operator for $\mathcal{A}[\theta], \theta \in \Theta$:

$$\mathcal{A}[\theta] = \sum_{k=0}^{K} A_k(t; \theta) T^{k-k_0},$$

where $T$ is the left-shift operator defined by:

$$(T x)(t) := x(t + 1) \text{ for } t \in \mathbb{Z},$$

and $k_0, K \in \mathbb{N}$. In order to be able to apply the shift operator $T$ and its inverse to $x \in \ell_2(\mathbb{N}_0)$, we first extend $x$ to $\tilde{x}$, defined by $\tilde{x}(t) := x(t)$, for $t \in \mathbb{N}_0$, and $\tilde{x}(t) = 0$, for $t < 0$. So we consider the elements of $\ell_2(\mathbb{N}_0)$ as elements of $\ell_2(\mathbb{Z})$ with zero past, that is, the causal part of $\ell_2(\mathbb{Z})$. For $A_k(t; \theta)$ we assume here the same decomposition property as in the continuous-time case:

$$A_k(t; \theta) = \sum_{j=1}^{J_k} B_{k,j}(\theta) C_{k,j}(t),$$

(3.12)

for some $J_k \in \mathbb{N}$, for each $k = 0, \ldots, K$. In a general context $J_k$ can be replaced by a constant independent of $k$. Substituting the expression (3.12) of $A_k$ in (3.8), we obtain:

$$Q(x; \theta) = \sum_{k=0}^{K} \sum_{j=1}^{J_k} B_{k,j}(\theta) \langle C_{k,j}(t) T^{k-k_0} x, x \rangle$$

$$= \sum_{k=0}^{K} \sum_{j=1}^{J_k} B_{k,j}(\theta) \mathcal{M}_{k,j}(x).$$
The expressions $\mathcal{M}_{k,j}(x) := \langle C_{k,j}^T x^k, x \rangle$ are independent of $\theta$ and completely determined if $x$ is known. The admissible space $\mathcal{D}_d$ for the discrete-time case is defined by:

$$\mathcal{D}_d = \left\{ x \in \mathcal{H} \mid \|C_{k,j}^T x^k, x\| < +\infty, \ k = 0, \ldots, K; \ j = 1, \ldots, J_k \right\}. \quad (3.13)$$

**Minimization Problem 3.1.4** For $x \in \mathcal{D}_d, x \neq 0$, find $\theta \in \Theta$ minimizing:

$$Q(x; \theta) = \sum_{k=0}^K \sum_{j=1}^{J_k} B_{k,j}(\theta) \mathcal{M}_{k,j}(x),$$

where $\theta$ is the compaction parameter, $\mathcal{M}_{k,j}(x)$ is a measurement on $x$ and $Q[\theta]$ is the enforced convergence criterion to be optimized. The criterion $Q[\theta]$ is also called the compaction criterion. The optimal value of $\theta$ is denoted by $\hat{\theta}$.

Assuming that the functions $B_{k,j}$ are differentially dependent on the parameter $\theta$, the candidates for the optimality are obtained from the boundary of $\Theta$ and:

$$\nabla_\theta Q[\theta] = \sum_{k=0}^K \sum_{j=1}^{J_k} \mathcal{M}_{k,j}(x) \nabla_\theta B_{k,j}(\theta) = 0, \quad (3.14)$$

for a fixed $x$ in the admissible space $\mathcal{D}_d$.

The kind of equations arising from (3.11) or (3.14) are easier to solve than the optimization of the quadratic error criterion $E_{\tau\tau}(N; x; \theta)$. The expressions (3.11) and (3.14) give candidates for local optimality in the interior of $\Theta$. For the optimality in $\Theta$, the boundaries of $\Theta$ have to be investigated separately, for instance, by Lagrange’s method. In general, these procedures result in a description of the optimal parameter explicitly as a function of $\mathcal{M}_{k,j}(x)$. We do not need to know $x$ completely, but only knowledge of the quantities $\mathcal{M}_{k,j}(x)$ is required; see subsequent sections. In contrast to optimal parameters according to the quadratic error criterion, the compaction criterion has the convenient property that it defines the parameter independent of the number of the terms in the expansion. This is a consequence of (3.3). By means of examples we will show that this compaction procedure gives results close to the optimal ones in the quadratic sense. If the results obtained by the enforced convergence criterion are not sufficiently accurate, then the optimal parameters obtained by the compaction criterion can always be used as initial estimate in a more refined optimization criterion.

**3.2 Second-order operator**

Most of the classical orthogonal functions are particular solutions of a second-order ordinary difference or differential equation. In the next two subsections, we consider
3.2 Second-order operator

<table>
<thead>
<tr>
<th>Type</th>
<th>Laguerre</th>
<th>Jacobi</th>
<th>Hermite</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{H}$</td>
<td>$L_2(0, \infty)$</td>
<td>$L_2(-1, 1)$</td>
<td>$L_2(\mathbb{R})$</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>${\alpha &gt; -1, \sigma &gt; 0}$</td>
<td>${\alpha &gt; -1, \beta &gt; -1}$</td>
<td>${\sigma &gt; 0, t_0 \in \mathbb{R}}$</td>
</tr>
<tr>
<td>$A_0$</td>
<td>$\frac{(\alpha - \sigma \beta)^2 - 2\sigma t}{\sigma \rho}$</td>
<td>$\frac{2\sigma^2 (\alpha + \beta)^2 - 1}{4}$</td>
<td>$\frac{\sigma^2 (t - t_0)^2 - 1}{2}$</td>
</tr>
<tr>
<td>$A_1$</td>
<td>$\frac{-1}{\sigma}$</td>
<td>$2t$</td>
<td>$0$</td>
</tr>
<tr>
<td>$A_2$</td>
<td>$\frac{-t}{\sigma}$</td>
<td>$t^2 - 1$</td>
<td>$-\frac{1}{2\sigma^2}$</td>
</tr>
<tr>
<td>$\theta$</td>
<td>$(\alpha, \beta)$</td>
<td>$n$</td>
<td>$n(n + \alpha + \beta + 1)$</td>
</tr>
<tr>
<td>$\eta_n$</td>
<td>$n$</td>
<td>$(\alpha, \beta)$</td>
<td>$(t_0, \sigma)$</td>
</tr>
</tbody>
</table>

Table 3.1: Differential operator coefficients $A_k(t; \theta)$ for Laguerre, Jacobi and Hermite functions.

this kind of equation. We define the problem of optimization for continuous- and discrete-time case. In Subsection 3.2.1, we deal with orthogonal expansions related to second-order differential equations. In Subsection 3.2.2, we deal with the case of second-order difference equations.

3.2.1 Second-order differential operator

Let $\{\phi_n[\theta]| n \in \mathbb{N}_0\}$ be a complete orthonormal system in a Hilbert space $\mathcal{H}$ corresponding to the second-order differential equation:

$$A_2(t; \theta)y'' + A_1(t; \theta)y' + (A_0(t; \theta) - \eta_n[\theta])y = 0,$$

(3.15)

where the coefficients $A_0(t; \theta), A_1(t; \theta),$ and $A_2(t; \theta)$ depend on the type of orthonormal basis $\{\phi_n[\theta]| n \in \mathbb{N}_0\}$ independent of the index $n$. The functions $A_k(t; \theta), k = 0, 1, 2,$ depend also on the free parameter $\theta$. As a special case we mention the Laguerre, Jacobi and Hermite differential operator, which are given in Table 3.1. These specific cases will be treated later in this part. Furthermore, the functions $A_k(t; \theta),$ for $k = 0, 1, 2,$ can be expressed as a summation of separable functions of $\theta$ and $t$:

$$A_k(t; \theta) = \sum_{j=1}^{J_k} B_{k,j}(\theta) C_{k,j}(t),$$

(3.16)

for some $J_k \in \mathbb{N}, k = 0, 1, 2$. For simplicity in notation, we write:

$$A_k(t; \theta) = B_k(\theta) C_k^t(t),$$

(3.17)
Table 3.2: Variable separation of the coefficients \( A_k(t; \theta) \) in the differential operator of Laguerre and Jacobi functions.

<table>
<thead>
<tr>
<th>Type</th>
<th>( \theta )</th>
<th>( J_0 )</th>
<th>( B_0(\theta) )</th>
<th>( C_0(t) )</th>
<th>( J_1 )</th>
<th>( B_1(\theta) )</th>
<th>( C_1(t) )</th>
<th>( J_2 )</th>
<th>( B_2(\theta) )</th>
<th>( C_2(t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta )</td>
<td>((\alpha, \sigma))</td>
<td>((\alpha, \beta))</td>
<td>(3)</td>
<td>(\left(\frac{\alpha}{2}, \frac{\beta}{2}, -\frac{1}{2}(\alpha + 1)\right))</td>
<td>(t, (t^2, 1))</td>
<td>(1)</td>
<td>(-\frac{1}{\sigma})</td>
<td>(1)</td>
<td>(t)</td>
<td>(t^2 - 1)</td>
</tr>
<tr>
<td>( B_0(\theta) )</td>
<td>(\left(\frac{\alpha}{2}, \frac{\beta}{2}, -\frac{1}{2}(\alpha + 1)\right))</td>
<td>(3)</td>
<td>(\left(\frac{\alpha^2 + \beta^2}{2}, -\frac{1}{2}(\alpha + \beta + 1)^2 - 1\right))</td>
<td>(1)</td>
<td>(2t)</td>
<td>(1)</td>
<td>(-\frac{1}{\sigma})</td>
<td>(1)</td>
<td>(t^2 - 1)</td>
<td></td>
</tr>
</tbody>
</table>

where:

\[
B_k(\theta) := (B_{k,1}(\theta), B_{k,2}(\theta), \ldots, B_{k,J_k}(\theta))
\]

\[
C_k(t) := (C_{k,1}(t), C_{k,2}(t), \ldots, C_{k,J_k}(t))
\]

for \( J_0, J_1, J_2 \in \mathbb{N} \). In Table 3.2 we represent the description of the coefficients \( A_0, A_1 \) and \( A_2 \), in the case of Laguerre, Jacobi and Hermite differential operator. We note that the separation of variables (3.17) is not unique. So several decomposition solutions can be obtained. For example in the case of the Jacobi differential operator we can choose \(\left(\frac{\alpha^2 + \beta^2}{2}, -\frac{1}{2}(\alpha + \beta + 1)^2 - 1\right)\) and \(\left(t^2, 1\right)\) for \(B_0\) and \(C_0\), respectively, instead of the solution given in Table 3.2. Also \(J_0, J_1, J_2\), are not unique. An example can be obtained from Jacobi case, where \(J_2 = 2\), \(B_2(\theta) = (1, -1)\) and \(C_2(t) = (t^2, 1)\) is also a solution for variable separation of \(A_2(t; \theta)\). In this way, the measurements \(\mathcal{M}_{k,j}(x)\) change, which may have a consequence on the simplicity of the admissible space \(\mathcal{D}\). Sometimes, manipulations in \(C_{k,j}\) are useful to simplify the expression of \(Q(x; \theta)\), and consequently the admissible space \(\mathcal{D}\). The optimality problem according to the enforced convergence criterion can be stated as follows:
3.2 Second-order operator

Minimization Problem 3.2.1  For $x \in \mathcal{D}_c$. Minimize

$$Q(x; \theta) = \sum_{k=0}^{2} \sum_{j=1}^{J} \langle C_{k,j}(t) \left( \frac{d}{dt} \right)^k x, x \rangle B_{k,j}(\theta),$$

(3.18)

with respect to $\theta \in \Theta$.

In the Jacobi case, Table 3.2, the expression of $Q(x; \theta)$ becomes:

$$Q(x; \theta) = \sum_{k=0}^{2} B_k(\theta) \langle C_x(t) \left( \frac{d}{dt} \right)^k x, x \rangle$$

$$= \frac{\alpha^2}{2} \left( \frac{1}{\sqrt{1-t^2}} x, \frac{1}{\sqrt{1-t^2}} x \right) + \frac{\beta^2}{2} \left( \frac{1}{\sqrt{1+t^2}} x, \frac{1}{\sqrt{1+t^2}} x \right) -$$

$$\frac{(\alpha + \beta + 1)^2}{4} \langle x, x \rangle + \langle 2t \frac{d}{dt} x, x \rangle + \langle (t^2 - 1) \left( \frac{d}{dt} \right)^2 x, x \rangle$$

$$= \frac{\alpha^2}{2} \left( \frac{1}{\sqrt{1-t^2}} x, \frac{1}{\sqrt{1-t^2}} x \right) + \frac{\beta^2}{2} \left( \frac{1}{\sqrt{1+t^2}} x, \frac{1}{\sqrt{1+t^2}} x \right) -$$

$$\frac{(\alpha + \beta + 1)^2}{4} \langle x, x \rangle + \langle (1 - t^2) \frac{d}{dt} x, \frac{d}{dt} x \rangle.$$  (3.19)

The functional $Q[\theta]$ is bounded in the admissible space $\mathcal{D}_c = \{x | x, \frac{d}{dt} x, \frac{1}{\sqrt{1-x^2}} x \in L_2(-1, 1)\}$, for $\theta \in \Theta$. As we remarked before, the expression $Q(x; \theta)$ defines an upper bound for the quadratic error $\mathcal{E}_{\text{err}}(N, x; \theta)$ by: $\mathcal{E}_{\text{err}}(N, x; \theta) \leq Q(x; \theta)/N^2$, for $N \in \mathbb{N}$. So if we want the quadratic error $\mathcal{E}_{\text{err}}(N, x; \theta)$ not to exceed $\epsilon$, then we have to choose for $N: N \geq \sqrt{Q(x; \theta)}/\epsilon$. In Section 3.3 we elaborate on this case. Assuming that the functions $B_{k,j}$ are differentially dependent on the parameter $\theta$, the candidates for local optimality in the interior of $\Theta$ are obtained from:

$$\nabla_{\theta} Q[\theta] = \sum_{k=0}^{2} \sum_{j=1}^{J} \mathcal{M}_{k,j}(x) \nabla_{\theta} B_{k,j}(\theta) = 0,$$  (3.20)

where the expressions $\mathcal{M}_{k,j}(x) := \langle C_{k,j}(t) \left( \frac{d}{dt} \right)^k x, x \rangle$ are completely determined if $x$ is known. For the optimality in $\Theta$ the boundaries must be taken into account.

3.2.2 Second-order difference operator

Similarly to the continuous case, let $\{\phi_n[\theta] | n \in \mathbb{N}_0\}$ be a complete orthonormal system in a Hilbert space $\mathcal{H}$ adhering to a difference equations of the second-order:

$$A_2(t; \theta) T y + (A_1(t; \theta) - \eta[\theta]) y + A_0(t; \theta) T^{-1} y = 0,$$  (3.21)
Optimal parameters in orthonormal series expansions

<table>
<thead>
<tr>
<th>Type</th>
<th>Charlier</th>
<th>Meixner</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{H}$</td>
<td>$\ell_2(\mathbb{N}_0)$</td>
<td>$\ell_2(\mathbb{N}_0)$</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>${\alpha \in \mathbb{R}}$</td>
<td>${</td>
</tr>
<tr>
<td>$A_0$</td>
<td>$-\alpha \sqrt{t}$</td>
<td>$-\xi \sqrt{(t+b-1)t}$ $\frac{1}{1-\xi^2}$</td>
</tr>
<tr>
<td>$A_1$</td>
<td>$t + \alpha^2$</td>
<td>$b\xi^2 + t(1+\xi^2)$ $\frac{1}{1-\xi^2}$</td>
</tr>
<tr>
<td>$A_2$</td>
<td>$-\alpha \sqrt{t+T}$</td>
<td>$-\xi \sqrt{(t+b)(t+1)}$ $\frac{1}{1-\xi^2}$</td>
</tr>
<tr>
<td>$\theta$</td>
<td>$\alpha$</td>
<td>$\xi$</td>
</tr>
<tr>
<td>$\eta_n$</td>
<td>$n$</td>
<td>$n$</td>
</tr>
</tbody>
</table>

Table 3.3: Difference operator $A_k(t; \theta)$ for Charlier and Meixner functions.

where $A_0[\theta], A_1[\theta], \text{and} A_2[\theta]$ depend on the type of functions considered, independently of $n$. The functions $A_k(t; \theta), k = 0, 1, 2,$ depend also on the free parameter. The cases of Charlier and Meixner functions are given in Table 3.3. These cases will be elaborated in Section 3.4. We note that $b$ in the Meixner case is not taken as parameter. This is due to the fact that $b$ cannot be separated from $t$ in a finite sum.

Furthermore, in the same way as for the continuous-time case, the functions $A_k(t; \theta), k = 0, 1, 2,$ can be expressed as a summation of separable functions of $\theta$ and $t$:

$$A_k(t; \theta) = \sum_{j=1}^{J_k} B_{k,j}(\theta) C_{k,j}(t), \quad (3.22)$$

for some $J_0, J_1, J_2 \in \mathbb{N}$. For simplicity in notation, we write:

$$A_k(t; \theta) = B_k(\theta) C_k(t), \quad k = 0, 1, 2, \quad (3.23)$$

where:

$$B_k(\theta) = (B_{k,1}(\theta), B_{k,2}(\theta), \ldots, B_{k,J_k}(\theta))$$

$$C_k(t) = (C_{k,1}(t), C_{k,2}(t), \ldots, C_{k,J_k}(t)),$$

for some $J_0, J_1, J_2 \in \mathbb{N}$. The enforced optimality criterion for the second-order difference operator is:
Minimization Problem 3.2.2  Let \( x \in \mathcal{D} \). Minimize

\[
Q[\theta] = \sum_{k=0}^{2} \sum_{j=1}^{J_k} \langle C_{k,j}(t) T^{k-1} x, x \rangle B_{k,j}(\theta),
\]

with respect to \( \theta \in \Theta \).

In Section 3.4, we illustrate the discrete-time case by Charlier and Meixner systems.

### 3.3 Optimal parameters in Jacobi expansions

The Jacobi expansions can be used to approximate functions with bounded support. The Jacobi functions constitute an orthonormal basis in \( L_2(-1, 1) \). They are well-known in harmonic analysis for their symmetric properties and their relationships with trigonometric functions. In this section we elaborate further on the Jacobi system. We compute the explicit formulas for the optimal parameters and properties in trigonometric functions. Moreover, we give the expression of the optimal parameter of the special case of the Jacobi system, namely, the Gegenbauer system and conclude the section by an example. We start with the definition of the orthonormal Jacobi functions:

**Definition 3.3.1** The Jacobi functions are defined by

\[
\mathcal{J}_n(t; \alpha, \beta) = h_n[\theta] \sum_{k=0}^{n} \binom{n}{k} \frac{\Gamma(n+\alpha+\beta+k+1)}{\Gamma(\alpha+k+1)} \left( \frac{-1}{t} \right)^k \left( 1 - t \right)^{k+n+\alpha} \left( 1 + t \right)^{\beta},
\]

where

\[
h_n[\theta] = \frac{\sqrt{(2n+\alpha+\beta+1)} \Gamma(n+\alpha+1) \Gamma(n+\alpha+\beta+1)}{\sqrt{n!} \Gamma(n+\beta+1) \Gamma(n+\alpha+\beta+1)} 2^{-n+\frac{\alpha+\beta+1}{2}},
\]

for \( n \in \mathbb{N}_0 \).

We recall the expression of the enforced convergence criterion for the Jacobi operator (3.19):

\[
Q(x; \theta) = \frac{\alpha^2}{2} \langle \frac{1}{\sqrt{1-t}} x, \frac{1}{\sqrt{1-t}} x \rangle + \frac{\beta^2}{2} \langle \frac{1}{\sqrt{1+t}} x, \frac{1}{\sqrt{1+t}} x \rangle - \frac{(\alpha + \beta + 1)^2 - 1}{4} \langle x, x \rangle + \langle (1 - t^2) \frac{d}{dt} x, \frac{d}{dt} x \rangle.
\]
Let \( x \in D_c = \{ x| x, \frac{d}{dx}x, \frac{1}{\sqrt{1-t^2}}x \in L_2(-1, 1) \} \) be fixed, \( x \neq 0 \). Further we consider the coefficients of \( \alpha^2 \) and \( \beta^2 \):

\[
Q(x; \theta) = \frac{\alpha^2}{4} \langle \frac{1-t}{1-t}x, x \rangle + \frac{\beta^2}{4} \langle \frac{1-t}{1-t}x, x \rangle - \frac{(\alpha + 1)(\beta + 1) - 1}{2} \langle x, x \rangle + (1 - t^2) \langle \frac{d}{dx}x, \frac{d}{dx}x \rangle.
\]

Since \( \langle \frac{1-t}{1-t}x, x \rangle > 0 \) and \( \langle \frac{1-t}{1-t}x, x \rangle > 0 \), for all \( x \in D_c^* \) and since

\[
\left| \begin{array}{cc}
\langle \frac{1-t}{1-t}x, x \rangle & -\langle x, x \rangle \\
-\langle x, x \rangle & \langle \frac{1-t}{1-t}x, x \rangle
\end{array} \right| > 0,
\]

(3.24)

Thus the expression \( Q(x; \theta) \), for a fixed \( x \in D_c^* \) represents an ellipsoid, as a function of \((\alpha, \beta)\). So the expression \( Q[\theta] \) achieves the minimum at the center \( \theta_c \) of this ellipsoid, if the center \( \theta_c = (\alpha_c, \beta_c) \) belongs to \( \Theta = \{ (\alpha, \beta) \in \mathbb{R}^2 | \alpha > -1, \beta < -1 \} \). We note that the center \( \theta_c \) can be obtained from:

\[
\nabla_{\theta} Q[\theta_c] = \left( \frac{\partial}{\partial \alpha} Q[\theta_c], \frac{\partial}{\partial \beta} Q[\theta_c] \right) = 0.
\]

The partial derivatives of \( Q[\theta] \) are given by:

\[
\frac{\partial}{\partial \alpha} Q[\theta] = \frac{1}{2} \alpha \langle 1 + t, x \rangle - \frac{1}{2} (\beta + 1) \langle x, x \rangle,
\]

\[
\frac{\partial}{\partial \beta} Q[\theta] = \frac{1}{2} \beta \langle 1 - t, x \rangle - \frac{1}{2} (\alpha + 1) \langle x, x \rangle.
\]

Since the determinant in (3.24) is positive for \( x \in D_c^* \), the minimum of \( Q[\theta] \), is achieved by:

\[
\left( \begin{array}{c}
\hat{\alpha} \\
\hat{\beta}
\end{array} \right) = \left( \begin{array}{cc}
\langle \frac{1-t}{1-t}x, x \rangle & -\langle x, x \rangle \\
-\langle x, x \rangle & \langle \frac{1-t}{1-t}x, x \rangle
\end{array} \right)^{-1} \left( \begin{array}{c}
\langle x, x \rangle \\
\langle x, x \rangle
\end{array} \right).
\]

**Theorem 3.3.2** For \( x \) in the admissible space \( D_c, x \neq 0 \), the optimum compaction parameters \( \theta = (\hat{\alpha}, \hat{\beta}) \) for the Jacobi expansion are given by:

\[
\hat{\alpha} = \frac{2 \langle x, x \rangle \langle \frac{1-t}{1-t}x, x \rangle}{\langle \frac{1-t}{1-t}x, x \rangle^2 \langle \frac{1-t}{1-t}x, x \rangle - \langle x, x \rangle^2},
\]

(3.25)

\[
\hat{\beta} = \frac{2 \langle x, x \rangle \langle \frac{1-t}{1-t}x, x \rangle}{\langle \frac{1-t}{1-t}x, x \rangle^2 \langle \frac{1-t}{1-t}x, x \rangle - \langle x, x \rangle^2}.
\]

(3.26)
3.3 Optimal parameters in Jacobi expansions

We note that the expressions for $\hat{\alpha}$ and $\hat{\beta}$ are symmetrical:

$$\langle \frac{1}{\tau^2} x, x \rangle \hat{\alpha} = \langle \frac{1}{\tau^2} x, x \rangle \hat{\beta}.$$ 

The substitution of the optimal parameters $\hat{\alpha}$ and $\hat{\beta}$ into $Q(x; \theta)$ leads to

$$Q(x; \theta) = -\frac{1}{4} \langle x, x \rangle (\hat{\alpha} + \hat{\beta}) + \langle 1 - t^2 \frac{d}{dt} x, \frac{d}{dt} x \rangle.$$ 

Mathematically, Jacobi functions are known in the literature for their properties related to trigonometric functions. To transform the Jacobi functions to the trigonometric domain, we substitute $t = -\cos 2\tau$, $\tau \in (0, \pi/2)$. This way we transform $L_2(-1, 1)$ onto $L_2(0, \pi/2)$. Then we have for the principal terms in the Jacobi functions:

$$(1 - t)^{k+\alpha/2} (1 + t)^{\beta/2} = 2^{k+(\alpha+\beta)/2} (\cos \tau)^{2k+\alpha} (\sin \tau)^{\beta}.$$ 

For simplicity in notation we define $x_c(\tau) = x(-\cos 2\tau)$, $\tau \in (0, \pi/2)$. For the compaction criterion and the optimal parameters, as functions of $x_c$ we have

$$Q(x_c, \theta) = \alpha^2 \langle \sin^2 \tau \tan \tau \cdot x_c, x_c \rangle + \beta^2 \langle \cos^2 \tau \cot \tau \cdot x_c, x_c \rangle -

((\alpha + 1)(\beta + 1) - 1) \langle \sin 2\tau \cdot x_c, x_c \rangle + \frac{1}{2} \langle \sin 2\tau \cdot \frac{dx_c}{d\tau}, \frac{dx_c}{d\tau} \rangle, \quad (3.27)$$

$$\hat{\alpha} = \frac{2\langle \sin 2\tau \cdot x_c, x_c \rangle \langle \cot \tau \cdot x_c, x_c \rangle}{4\langle \cos^2 \tau \cot \tau \cdot x_c, x_c \rangle \langle \sin^2 \tau \tan \tau \cdot x_c, x_c \rangle - \langle \sin 2\tau \cdot x_c, x_c \rangle^2},$$

$$\hat{\beta} = \frac{2\langle \sin 2\tau \cdot x_c, x_c \rangle \langle \tan \tau \cdot x_c, x_c \rangle}{4\langle \cos^2 \tau \cot \tau \cdot x_c, x_c \rangle \langle \sin^2 \tau \tan \tau \cdot x_c, x_c \rangle - \langle \sin 2\tau \cdot x_c, x_c \rangle^2},$$

where $x \in D_c$, $x_c \in D_{cc}$:

$$D_{cc} = \left\{ x_c | x_c, \frac{1}{\sqrt{\sin 2\tau}} x_c, \frac{1}{\sqrt{\sin 2\tau}} \frac{dx_c}{d\tau} \in L_2(0, \pi/2) \right\}.$$

The optimal parameter in orthonormal series expansions

\[
\langle x, x \rangle = 2\langle \sin 2\tau \cdot x_c, x_c \rangle,
\]
\[
\langle \frac{1}{1+t} x, x \rangle = 2\langle \cot \tau \cdot x_c, x_c \rangle,
\]
\[
\langle \frac{1}{1-t} x, x \rangle = 2\langle \tan \tau \cdot x_c, x_c \rangle,
\]
\[
\langle \frac{1-t}{1+t} x, x \rangle = 4\langle \cos^2 \tau \cot \tau \cdot x_c, x_c \rangle,
\]
\[
\langle \frac{1+t}{1-t} x, x \rangle = 4\langle \sin^2 \tau \tan \tau \cdot x_c, x_c \rangle,
\]
\[
\langle (1-t^2) \frac{dx}{dt}, \frac{dx}{dt} \rangle = \frac{1}{2} \langle \sin 2\tau \cdot \frac{dx_c}{d\tau}, \frac{dx_c}{d\tau} \rangle.
\]

The minimum compaction criterion in terms of \( x_c \) and trigonometric functions becomes

\[
Q(x_c, \theta) = -\frac{1}{2} \langle \sin 2\tau \cdot x_c, x_c \rangle (\hat{\alpha} + \hat{\beta}) + \frac{1}{2} \langle \sin 2\tau \cdot \frac{dx_c}{d\tau}, \frac{dx_c}{d\tau} \rangle.
\]

Gegenbauer, Chebyshev and Legendre functions are special cases of the Jacobi functions, obtained by \( \beta = \alpha \). The Gegenbauer functions \( C_n[\lambda] \) are special cases of the Jacobi functions given by the parameter \( \theta = (\lambda - 1/2, \lambda - 1/2) \):

\[
C_n(t; \lambda) = J_n(t; \lambda - \frac{1}{2}, \lambda - \frac{1}{2}).
\]  

(3.28)

The Chebyshev functions \( T_n, U_n \) and the Legendre functions \( P_n \) are special cases of the Gegenbauer functions given by \( \lambda = 0, 1, 1/2 \):

\[
T_n(t) = C_n(t; 0),
\]
\[
U_n(t) = C_n(t; 1),
\]
\[
P_n(t) = C_n(t; \frac{1}{2}).
\]

The compaction criterion in the case \( \beta = \alpha \) becomes:

\[
Q(x_c; \theta) = 2\alpha^2 \langle \frac{\cos^2 2\tau}{\sin 2\tau} x_c, x_c \rangle - 2\alpha \langle \sin 2\tau \cdot x_c, x_c \rangle + \frac{1}{2} \langle \sin 2\tau \cdot \frac{dx_c}{d\tau}, \frac{dx_c}{d\tau} \rangle.
\]

The optimal parameter \( \hat{\alpha} \) in this case is

\[
\hat{\alpha} = \frac{1}{2} \frac{\langle \sin 2\tau \cdot x_c, x_c \rangle}{\langle \frac{\cos^2 2\tau}{\sin 2\tau} x_c, x_c \rangle}.
\]
3.3 Optimal parameters in Jacobi expansions

The minimum compaction criterion in the case $\beta = \alpha$ is:

$$Q(x_e, \hat{\theta}) = -\langle \sin 2\tau \cdot x_e, x_e \rangle \alpha + \frac{1}{2} \langle \sin 2\tau \cdot \frac{dx_e}{d\tau}, \frac{dx_e}{d\tau} \rangle.$$  

So the optimal parameter $\hat{\lambda}$ for the Gegenbauer series expansions, for $x \in \mathcal{D}_c^*$, is

$$\hat{\lambda} = \alpha + \frac{1}{2} = \frac{1}{2} \langle \sin 2\tau \cdot x_e, x_e \rangle.$$

For the Chebyshev functions, $U_n = C_n[1]$, as defined in (3.28), we have no free parameters. However, in this case it is more interesting to know when we can use the Chebyshev system to approximate a function. The question is now formulated in an opposite way: what are the functions $x_e$ such that:

$$\alpha = \frac{1}{2} \langle \sin 2\tau \cdot x_e, x_e \rangle = 1. \quad (3.29)$$

It is desired to use Chebyshev expansions for the function $x_e$ if $x_e$ satisfies the following integral equation:

$$\int_0^{\pi} \frac{\cos 4\tau}{\sin 2\tau} \left\{ |x_e(\tau)|^2 + |x_e\left(\frac{\pi}{2} - \tau\right)|^2 \right\} d\tau = 0.$$

A sufficient condition for the optimality of Chebyshev expansions is

$$|x_e(\tau)|^2 + |x_e\left(\frac{\pi}{2} - \tau\right)|^2 = \sin 2\tau. \quad (3.30)$$

Examples satisfying this equation are $\sqrt{\sin \tau \cos \tau}$, $\sqrt{\sin 2\tau \sin \tau}$ and $\sqrt{\sin 2\tau \cos \tau}$. It is possible to give an explicit construction of $x_e$ satisfying (3.30): let $u$ be a function such that the function $u^2$ is symmetric with respect to $\tau = \frac{\pi}{4}$ with support($u$) $\subseteq [0, \frac{\pi}{2}]$ and

$$\int_0^{\pi} u^2(\tau) d\tau = \frac{\pi}{2}.$$

Define the function

$$u(\tau) = \int_0^{\tau} u^2(\lambda) d\lambda, \ \tau \in [0, \frac{\pi}{2}].$$

Since $u^2$ is symmetric with respect to $\tau = \frac{\pi}{4}$ we have $u(\tau) + u\left(\frac{\pi}{2} - \tau\right) = \frac{\pi}{2}$. Now we can put

$$x_e(\tau) = \sqrt{\sin 2\tau \sin u(\tau)},$$

as solution for (3.30). This way we form an idea about the functions for which it is desired to use Chebyshev expansions.
Example 3.3.3 As example we consider the function $x(t) = (1 - t)^p (1 + t)^q$, $t \in (-1,1)$, $p, q > \frac{1}{2}$. For this example we can obviously establish that the optimal parameters for $\alpha$ and $\beta$ are $2p$ and $2q$. But the question is: what are the optimal parameters according to the compaction criterion? We consider the same substitution as before, namely $t = -\cos 2\tau$, $\tau \in (0, \pi/2)$. Then we have

$$x_c(\tau) = 2^{p+q} \cos^{2p} \tau \sin^{2q} \tau.$$ 

The optimal parameters according to the compaction criterion are:

$$\hat{\alpha} = 2p,$$
$$\hat{\beta} = 2q,$$

since

$$\langle \sin 2\tau \cdot x_c, x_c \rangle = 2^{2(p+q)} B(2q + 1, 2p + 1),$$
$$\langle \cot \tau \cdot x_c, x_c \rangle = 2^{2(p+q)-1} B(2q, 2p + 1),$$
$$\langle \tan \tau \cdot x_c, x_c \rangle = 2^{2(p+q)-1} B(2q + 1, 2p),$$
$$\langle \cos^2 \tau \cot \tau \cdot x_c, x_c \rangle = 2^{2(p+q)-1} B(2q, 2p + 2),$$
$$\langle \sin^2 \tau \tan \tau \cdot x_c, x_c \rangle = 2^{2(p+q)-1} B(2q + 2, 2p),$$

where $B$ is the beta function:

$$B(q, p) = \int_0^1 t^{q-1}(1-t)^{p-1} dt.$$ 

Since $x$, in this case, is represented in the first Jacobi function, we have:

$$Q(x_c; \hat{\theta}) = 0.$$ 

3.4 Charlier and Meixner optimal parameters

The Charlier and Meixner functions define orthonormal bases in $\ell_2(\mathbb{N}_0)$. In this section we compute the optimal parameter for each of these systems.

Definition 3.4.1 In $\ell_2(\mathbb{N}_0)$ we define the following basis functions:
1. The Charlier functions are defined by:

\[ \phi_n(t; \theta) = [-|\alpha|]_n^2 e^{-|\alpha|^2/2} \alpha^t \sqrt{\frac{t!}{n!}} \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{\alpha^{-2k}}{(t-k)!}, \]

where \( \theta = \alpha \in \mathbb{R} \), is the Charlier parameter and \( t, n \in \mathbb{N}_0 \).

2. The Meixner functions are defined by:

\[ \phi_n(t; \theta, b) = \xi^{t+n} \sqrt{\frac{(1-\xi^2)b}{\Gamma(n+b)\Gamma(t+b)}} \sum_{k=0}^{n} (-1)^{k+n} \binom{n}{k} \frac{\xi^{-2k} \Gamma(t-k+n+b)}{(t-k)!}, \]

where \( \xi \in (-1, 1) \), \( \theta = \xi \), is the Meixner parameter and \( t, n \in \mathbb{N}_0 \), \( b \in \mathbb{R}^+ \).

The admissible spaces for Charlier \( \mathcal{D}_{dc} \) and Meixner \( \mathcal{D}_{dm} \) are identical and given by \( \mathcal{D}_d \):

\[ \mathcal{D}_d = \{ x | (1 + \sqrt{t})x \in \ell_2(\mathbb{N}_0) \}. \]

<table>
<thead>
<tr>
<th>Type</th>
<th>Charlier</th>
<th>Meixner</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta )</td>
<td>( \alpha )</td>
<td>( \xi )</td>
</tr>
<tr>
<td>( J_0 )</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( B_0(\theta) )</td>
<td>(-\alpha)</td>
<td>(-\frac{\xi}{1-\xi^2})</td>
</tr>
<tr>
<td>( C_0(t) )</td>
<td>( \sqrt{t} )</td>
<td>( \sqrt{(t+b-1)t} )</td>
</tr>
<tr>
<td>( J_1 )</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>( B_1(\theta) )</td>
<td>( (\alpha^2, 1) )</td>
<td>( (\frac{\xi^2}{1-\xi^2}, \frac{1+\xi^2}{1-\xi^2}) )</td>
</tr>
<tr>
<td>( C_1(t) )</td>
<td>( (1, t) )</td>
<td>( (b, t) )</td>
</tr>
<tr>
<td>( J_2 )</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( B_2(\theta) )</td>
<td>(-\alpha)</td>
<td>(-\frac{\xi}{1-\xi^2})</td>
</tr>
<tr>
<td>( C_2(t) )</td>
<td>( \sqrt{t+1} )</td>
<td>( \sqrt{(t+b)(t+1)} )</td>
</tr>
</tbody>
</table>

*Table 3.4: Separation of variables in the coefficients \( A_k(t; \theta) \) of the differential operator corresponding to Charlier and Meixner functions.*
The enforced convergence criteria \( Q(x; \theta) \) and \( Q(x; \theta, b) \) for Charlier and Meixner operators are respectively:

\[
Q(x; \theta) = \alpha^2 \langle x, x \rangle - 2\alpha \Re \{ \sqrt{1 + \xi x} \} + \langle \sqrt{i x}, \sqrt{i x} \rangle, \tag{3.31}
\]

\[
Q(x; \theta, b) = \frac{1}{1 - \xi^2} \left[ -2\xi \Re \{ \sqrt{1 + \xi x}, \sqrt{i + bx} \} + b \langle x, x \rangle + 2\langle \sqrt{i x}, \sqrt{i x} \rangle \right] - b \langle x, x \rangle - \langle \sqrt{i x}, \sqrt{i x} \rangle. \tag{3.32}
\]

We note that the parameter \( b \) is not included in \( \theta \), since it can not be separated from the measurements on \( x \). The separation coefficients of \( Q(x; \theta) \) and \( Q(x; \theta, b) \) are given Table 3.4.

**Theorem 3.4.2** For \( x \in D_d^* \),

1. the Charlier criterion \( Q(x; \theta) \) achieves a minimum at \( \hat{\theta} = \hat{\alpha} \):

\[
\hat{\alpha} = \frac{\Re \{ \sqrt{1 + \xi x} \} \langle x, x \rangle}{\langle x, x \rangle},
\]

and the minimum equals

\[
Q(x, \hat{\theta}) = \frac{\langle \sqrt{i x}, \sqrt{i x} \rangle}{\langle x, x \rangle} - \left( \frac{\Re \{ \sqrt{1 + \xi x} \} \langle x, x \rangle}{\langle x, x \rangle} \right)^2.
\]

2. the Meixner criterion, \( Q(x; \theta, b) \), is minimal at \( \hat{\theta} = \hat{\xi} \):

\[
\hat{\xi} = \beta(1 - \sqrt{1 - \frac{1}{\beta^2}}), \tag{3.33}
\]

where

\[
\beta = \frac{\langle \sqrt{i x}, \sqrt{i x} \rangle + b \langle x, x \rangle / 2}{\Re \{ \langle \sqrt{1 + \xi x}, \sqrt{i + bx} \rangle \}}, \tag{3.34}
\]

and the minimum of \( Q(x; \theta, b) \) equals

\[
Q(x; \hat{\theta}, b) = \left| \frac{\Re \{ \langle \sqrt{1 + \xi x}, \sqrt{i + bx} \rangle \} \langle x, x \rangle}{\langle x, x \rangle} \right| \sqrt{\beta^2 - 1} - \frac{b}{2}.
\]

\( \square \)

**Proof:** Let \( x \in D_d^* \).
3.5 Discussion

1. The candidates for the optimality are the zeros of the derivative of $Q(x; \theta)$ with respect to $\alpha$. Differentiation of $Q(x; \theta)$ with respect to $\theta = \alpha$ leads to

$$
\frac{dQ}{d\alpha}(x; \theta) = 2\langle x, x \rangle \left( \alpha - \frac{\Re\{\langle \sqrt{1 + T}x, x \rangle \}}{\langle x, x \rangle} \right).
$$

So $\frac{dQ}{d\alpha}(x; \hat{\theta}) = 0$ and $\frac{d^2Q}{d\alpha^2}(x; \hat{\theta}) > 0$. Thus the enforced convergence criterion $Q(x; \theta)$ for the Charlier system achieves the minimum in $\hat{\alpha}$.

2. Differentiation of $Q(x; \theta, b)$ with respect to $\xi$ leads to

$$
\frac{dQ}{d\xi}(x; \theta, b) = -2\Re\{\langle \sqrt{1 + T}x, \sqrt{1 + b}x \rangle \} \frac{\xi^2 - 2\beta \xi + 1}{(1 - \xi^2)^2}.
$$

There exists one candidate for the optimality, that is, $\hat{\xi}$ in (3.33). It is obvious to prove that $\beta^2 - 1 > 0$:

$$
|\langle \sqrt{1 + T}x, \sqrt{1 + b}x \rangle|^2 \leq \langle \sqrt{1 + T}x, \sqrt{1 + T}x \rangle \langle \sqrt{1 + b}x, \sqrt{1 + b}x \rangle
$$

$$
\leq \langle \sqrt{1 + T}x, \sqrt{1 + T}x \rangle \left[ \langle \sqrt{1 + T}x, \sqrt{1 + T}x \rangle + b(x, x) \right]
$$

$$
\leq \left( \langle \sqrt{1 + T}x, \sqrt{1 + T}x \rangle + b(x, x)/2 \right)^2.
$$

Furthermore,

$$
\frac{d^2Q}{d\xi^2}(x; \theta, b) = 8\Re\{\langle \sqrt{1 + T}x, \sqrt{1 + b}x \rangle \} (\beta^2 - 1) \frac{\hat{\xi} (1 - \hat{\xi})^3}{(1 - \xi^2)^2} > 0,
$$

where $\text{sgn}(\hat{\xi}) = \text{sgn}(\beta) = \text{sgn}(\Re\{\langle \sqrt{1 + T}x, \sqrt{1 + b}x \rangle \})$.

We note that if $\alpha, \xi \in \mathbb{C}$; with $|\xi| < 1$, then we can obtain similar results.

### 3.5 Discussion

We considered an enforced convergence criterion to establish the optimal parameters in series expansions using properties of classical orthonormal systems, which is an adapted version of the criterion defined in [20]. We gave a well-defined problem to determine the optimal parameters according to the (new) enforced convergence criterion.
We have shown that this enforced convergence criterion leads to simple expressions for the optimal parameters. Further, we gave an upper bound for the quadratic truncation error. The upper bound may become tighter if we take for the difference/differential operator powers of $A[ \theta ]$. This is not without consequences for the simplicity of the optimization enforced convergence criterion.

The optimal solution does not require complete knowledge of the function considered, but only some experimental measurements. This gives us opportunities for design purposes, that is, by straightforward numerical calculations on a measured impulse response we can give a condensed representation of the impulse response. In Chapter 4 we consider signal analysis by Hermite transformations, where the resulting optimum parameters have simple physical interpretations and the minimum enforced convergence criterion directly is related to time-bandwidth product. In Chapter 5 we consider system identification of causal systems by Kautz series expansions.

In contrast to the quadratic error criterion, the enforced convergence criterion has the convenient property that it defines the parameter independent of the number of the terms in the expansion. By means of an example we have shown that the enforced convergence criterion gives the expected suitable result.

Since the enforced convergence criterion allows for differential operators with eigenvalues dependent on the parameters, the optimal parameters for the Jacobi expansions can now be found. In this case, we gave the explicit expressions for the optimal parameters as functions of measurements on function in $L_2(−1, 1)$, we translated the expressions to trigonometric expressions and derived the optimal parameter for the Gegenbauer expansions. Further, we concluded the section with an academic example.

An interesting question could be: given a certain choice of parameters, what is the class of functions that has this parameters as optimal solution. We can treat this problem by putting the expressions for the optimal parameters equal to the values chosen. This way we can divide the admissible space into a collection of classes according to the values of the parameters. An illustration is given for the Chebyshev expansions, where the parameter is fixed.

The Charlier and Meixner expansions are examples for discrete-time orthogonal expansions in $l_2(\mathbb{N}_0)$. We have shown that the optimal parameters for the Charlier and Meixner systems are simple explicit expressions dependent on simple limited number of experimental data of the functions.
Chapter 4

Signal analysis by modulated Hermite transformations

The outline of this chapter is as follows. In Section 4.1 we will present the admissible space $\mathcal{D}_c$ of the considered functions and in Section 4.2 the harmonically modulated Hermite functions. In Section 4.3 we will look at some properties of the modulated Hermite functions, that is, the symmetry of these functions in their time and frequency description. In Section 4.4 we will consider a Hermite series expansion of an arbitrary function of the admissible space $\mathcal{D}_c$ and define a compaction criterion. Finally, we will calculate the optimum free parameters (translation, modulation, and scale) and consider their physical interpretation (Section 4.5). We conclude with a discussion.

The Fourier transformation is denoted by $\mathcal{F}$ and defined as

$$X(\omega) = \mathcal{F}\{x\}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt.$$ 

### 4.1 The considered class of signals

In this section we will introduce the class of functions in which we are interested.

**Definition 4.1.1** We define the admissible space $\mathcal{D}_c$, a subspace of $L_2(\mathbb{R})$, by

$$\mathcal{D}_c = \{x|tx, x' \in L_2(\mathbb{R})\}.$$ 

The reason for this choice of $\mathcal{D}_c$ is clarified in Section 4.3. In physical terms, the considered class of functions has finite energy and a finite second-order moment. The latter implies that each function within $\mathcal{D}_c$ has a finite effective support where the effective support is defined as the second-order central moment $\tilde{T}$ in the time domain

$$\tilde{T}^2 = \int_{-\infty}^{+\infty} |(t-T_1)x(t)|^2 dt / \langle x, x \rangle$$
with $T_1$ being the first-order moment, that is

$$T_1 = \int_{-\infty}^{+\infty} |x(t)|^2 \, dt.$$

Since $x' \in L^2(\mathbb{R})$, the Fourier transform $X(\omega)$ of $x$ has a finite effective support $\tilde{\Omega}$. This is defined as before

$$\tilde{\Omega}^2 = \int_{-\infty}^{+\infty} |(\omega - \Omega_1)X(\omega)|^2 \, d\omega / \langle x, x \rangle$$

with $\Omega_1$ being the first-order moment, that is

$$\Omega_1 = \int_{-\infty}^{+\infty} \omega |X(\omega)|^2 \, d\omega / \langle x, x \rangle.$$

Consequently, a function $x \in \mathcal{D}_c$ has a finite time-bandwidth product (TBWP) which is defined by

$$\text{TBWP} = \tilde{\Omega} T.$$

This product, as hypothesized by Gabor [43], is the number of degrees of freedom in a signal. In other words, for the class of functions having equal TBWP (and without any additional a priori information), the most compact transform would yield a number of parameters equal to the TBWP. This implies that the number of parameters or coefficients that needs to be transmitted or stored has as a lower bound this TBWP.

### 4.2 The modulated Hermite functions

**Definition 4.2.1** The modulated Hermite functions $\phi_n[\theta], n \in \mathbb{N}_0$, are defined by

$$\phi_0(t; \theta) = \frac{\sigma}{\sqrt{\pi}} e^{-\sigma^2 (t-T)^2/2 \theta} e^{i \theta t}$$

$$\phi_1(t; \theta) = \sqrt{2} \sigma (t - T) \phi_0(t; \theta),$$

and for $n \geq 2$

$$\phi_n(t; \theta) = \sigma (t - T) \frac{2}{n} \phi_{n-1}(t; \theta) - \sqrt{1 - \frac{1}{n}} \phi_{n-2}(t; \theta), \quad (4.1)$$

with $\theta = (\sigma, T, \Omega) \in \Theta = \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}$. \qed
4.3 Properties

The Hermite functions known from literature are different to the functions defined here. In the modulated Hermite functions we introduced a modulation factor, namely, $e^{j\omega t}$. Explicitly, \( \phi_n(t; \theta) = e^{j\omega t} H_n(t-T; \sigma) \) where \( H_n(t; \sigma) \) are the Hermite functions as defined in [20]. The Hermite functions are derived from the Hermite polynomials by \( H_n(t; \sigma) = \sqrt{\sigma} \sqrt{w(\sigma t)} P_n(\sigma t)/h_n \) where \( P_n(t) \) are the Hermite polynomials, \( w(t) = e^{-t^2} \) is the associated weight function, and \( h_n = \sqrt{2^n n!} \sqrt{\pi} \) is the normalization constant [20]. The basis functions can be easily generated using the recurrence relation, though this is not without numerical problems in the case of large \( n \).

**Lemma 4.2.2** The system \( \{ \phi_n[\theta]| n \in \mathbb{N}_0 \} \) of modulated Hermite functions constitutes an orthonormal basis in \( L_2(\mathbb{R}) \).

**Proof:** The system \( \{ H_n[\sigma]| n \in \mathbb{N}_0 \} \) is an orthonormal basis in \( L_2(\mathbb{R}) \) [20]. Further, \[ \langle \phi_m[\theta], \phi_n[\theta] \rangle = \langle H_m[\sigma], H_n[\sigma] \rangle = \delta_{m,n} \] where \( \delta_{m,n} \) is the Kronecker delta. The completeness of the system \( \{ \phi_n[\theta]| n \in \mathbb{N}_0 \} \) follows by contradiction. Suppose that the system is incomplete. Then there is a function \( x \in L_2(\mathbb{R}) \) with \( x \neq 0 \) such that \( \langle x, \phi_n[\theta] \rangle = 0 \) for all \( n \in \mathbb{N}_0 \). Consequently, we have that \( \langle x, H_n[\sigma] \rangle = 0 \) for \( n \in \mathbb{N}_0 \) with \( x(t) = x(t)e^{-j\omega t} \). Since \( x \in L_2(\mathbb{R}) \) and the system \( \{ H_n[\sigma]| n \in \mathbb{N}_0 \} \) is complete in \( L_2(\mathbb{R}) \), the assertion is contradicted.

We note that Lemma 4.2.2 is not a statement about the rate of convergence of such series. This is, however, a major issue in practical applications since it determines the compaction of the transform. Sections 4.4 and 4.5 deal with this issue.

4.3 Properties

There are many properties related to the modulated Hermite functions. These essentially follow from all the properties associated with the Hermite polynomials. We will consider the differential equation only, since this is directly related to the compaction criterion that will be introduced in the next section.

Furthermore, in anticipation of later results, we will consider the Fourier transforms of the modulated Hermite functions.

**Lemma 4.3.1** The modulated Hermite functions \( \phi_n[\theta] \) adhere to the following ordinary differential equation on \( \mathbb{R} \)

\[
\frac{-1}{2\sigma^2} \phi''_n[\theta] + \frac{j\Omega}{\sigma^2} \phi'_n[\theta] + \left[ \frac{\sigma^2(t-T)^2}{2} - \frac{1}{2} + \frac{\Omega^2}{2\sigma^2} \right] \phi_n[\theta] = n \phi_n[\theta]. \tag{4.2}
\]
Signal analysis by modulated Hermite transformations

Proof: This can be derived from the differential equation for the Hermite functions [20] and the relation \( \phi_n(t; \theta) = e^{\Omega t} H_n(t - T; \sigma) \). The differential equation (4.2) reduces to that defined in 3.1 for \( \Omega = 0 \).

Definition 4.3.2 The real functional \( Q[\theta] \) on \( D_c \) is defined as

\[
Q(x; \theta) = \frac{1}{2\sigma^2} \langle x', x' \rangle + \frac{\Omega}{\sigma^2} \Im \{ \langle x, x' \rangle \} + \left( \frac{\sigma^2}{2\sigma^2} - \frac{1}{2} \right) \langle x, x \rangle + \frac{\sigma^2}{2}(t - T)x, (t - T)x, \]

(4.3)

where \( \Im \) and \( \theta \) denote the imaginary part and the vector consisting of the parameters \( \sigma, \Omega, \) and \( T \), respectively.

If \( x \) is twice differentiable, then equation (4.3) can be rewritten to

\[
Q(x; \theta) = \left( \frac{1}{2\sigma^2} x'' + \frac{\sigma^2}{\sigma^2} x' \right) \left( \frac{\sigma^2(t - T)^2}{2} - \frac{1}{2} \right) \langle x, x \rangle, \]

(4.4)

This equation expresses the link between the introduced quadratic functional \( Q[\theta] \) and the modulated Hermite functions since, in this form, the functional \( Q[\theta] \) directly reflects part of the differential equation (4.2). Subsequently we have

\[
Q(\phi_n[\theta]; \theta) = n, \quad n \in \mathbb{N}_0.
\]

Furthermore, \( Q[\theta] \) is a positive, well-defined functional; see next section.

In the following lemma we need to express the dependence of the modulated Hermite functions \( \phi_n[\theta] \) and its Fourier transform \( \Phi_n[\theta] \) on the free parameters.

Lemma 4.3.3 The Fourier transforms of the modulated Hermite functions can be expressed in the modulated Hermite functions according to

\[
\Phi_n(\omega; \sigma, T, \Omega) = \sqrt{2\pi} (-j)^n \phi_n(\omega; 1/\sigma, \Omega, -T). \quad (4.5)
\]

Proof: The Hermite functions \( H_n[1] = \phi_n[1, 0, 0] \) are eigenfunctions of the Fourier transformation with eigenvalue \( \sqrt{2\pi} (-j)^n \); for example see [43, 88]. This can be adapted to the modulated Hermite functions \( \phi_n[\theta] \) using the properties of the Fourier transformation concerning scaling, translation, and modulation.

From the previous lemma it immediately follows that we have a recurrence relation and a differential equation for the Fourier transforms of the modulated Hermite functions.
Lemma 4.3.4 The recurrence relation for the Fourier transforms of the modulated Hermite functions is given by

\[ \sqrt{\frac{n+1}{2}} \Phi_{n+1}[\theta] = \frac{\psi(\omega - \Omega)}{\sigma} \Phi_n[\theta] + \sqrt{\frac{n}{2}} \Phi_{n-1}[\theta], \text{ for } n \in \mathbb{N}_0. \]

Proof: Take the recurrence relation (4.1) and use (4.5).

Lemma 4.3.5 The Fourier transforms of the modulated Hermite functions adhere to a second-order linear differential equation according to

\[ \frac{-\sigma^2}{2} \Phi_n''[\theta] + \frac{\sigma^2 T}{2} \Phi_n'[\theta] + \left[ \frac{(\omega - \Omega)^2}{2\sigma^2} - \frac{1}{2} + \frac{\sigma^2 T^2}{2} \right] \Phi_n[\theta] = n \Phi_n[\theta] \]

for \( n \in \mathbb{N}_0 \).

Proof: Take the differential equation (4.2) and use (4.5).

Using the property that the Fourier transformation is an orthogonal transformation, we can write the following:

Lemma 4.3.6 The system \( \{ \Phi_n[\theta] | n \in \mathbb{N}_0 \} \) of the Fourier transforms of the modulated Hermite functions constitute an orthonormal system in \( L^2(\mathbb{R}) \).

Note that the four previous lemmas show that there is a high degree of symmetry between the modulated Hermite functions and their Fourier transforms. The roles of the modulation and the translation are interchanged and the scale \( \sigma \) in the time domain becomes a scale \( 1/\sigma \) in the Fourier description.

4.4 Compaction criterion

Since the Hermite functions form an orthonormal basis in \( L^2(\mathbb{R}) \), every function in the admissible space \( D_c \) can be expressed in this basis. We will now define the (obvious) modulated Hermite transformation.

Definition 4.4.1 We define the modulated Hermite transformation \( \mathcal{Y} : L^2(\mathbb{R}) \rightarrow \ell^2(\mathbb{N}_0) \) by

\[ \mathcal{Y} \{ x \}(n) = \langle x, \phi_n[\theta] \rangle, \]

for \( x \in L^2(\mathbb{R}) \).
We can call $\mathcal{Y}\{x\}$ the expansion coefficients, the Fourier-Hermite spectrum (or simply Hermite spectrum), or the Hermite transform.

In practical situations, we use a truncated Hermite series expansion. This then constitutes an approximation of the expanded function. For reasons of coding, pattern recognition, and manipulation, we are interested in compact transforms. We introduce the following compaction measure.

**Definition 4.4.2** We define the compaction measure $\zeta(\theta)$ by the positive-valued measure

$$
\zeta(x; \theta) = \frac{\sum_{n \in \mathbb{K}} n|\mathcal{Y}\{x\}(n)|^2}{\|x\|^2}, \quad \text{for } x \in \mathcal{D}_c.
$$

□

The physical interpretation of this measure is obvious: it is the first-order moment of the energy distribution in the transform domain.

**Lemma 4.4.3** The compaction measure is directly related to the previously introduced functional, namely

$$
\zeta(x; \theta) = \frac{Q(x; \theta)}{\|x\|^2}.
$$

□

**Proof:** Substitute (4.4), develop $x$ in a Hermite series, and use (4.2).

This relation shows that the functional $Q[\theta]$ is a positive functional. We also note that this compaction measure can be translated into an upper bound for the quadratic error in a truncated Hermite expansion [20, 87], according to

$$
\sum_{n=N}^{\infty} |\mathcal{Y}\{x\}(n)|^2 \leq \frac{\zeta(x; \theta)}{N} \|x\|^2, \quad \text{for } N \geq \zeta(x; \theta).
$$

Though it may not be a very tight bound, it may be quite useful in practice, since it gives an indication of the required number of terms prior to the execution of the expansion.

### 4.5 Optimum compaction parameters

The Hermite transformation has three free variables ($T$, $\sigma$, and $\Omega$) which can be used to obtain a compact transform.
4.5 Optimum compaction parameters

**Definition 4.5.1** We define the optimum compaction parameters \( \hat{T}, \hat{\sigma}, \text{ and } \hat{\Omega} \) for a given function \( x \in \mathcal{D}_e \) as those \( T, \sigma, \text{ and } \Omega \) yielding the minimum compaction measure.

**Theorem 4.5.2** For a function \( x \in \mathcal{D}_e \) the optimum compaction parameters are given by

\[
\hat{T} = \frac{\langle tx, x \rangle}{\langle x, x \rangle} \quad (4.6)
\]
\[
\hat{\Omega} = -\frac{g(x', x)}{\langle x, x \rangle} \quad (4.7)
\]
\[
\hat{\sigma}^4 = \frac{\langle x', x' \rangle + \langle x', x \rangle^2 / \langle x, x \rangle}{\langle (t - \hat{T})x, (t - \hat{T})x \rangle}. \quad (4.8)
\]

**Proof:** Differentiation of equation (4.3) with respect to \( \Omega, T, \text{ and } \sigma \) and equating this to zero leads to the above results. Considering the second-order derivatives, it follows that these parameters yield the minimum of \( \zeta(x; \theta) \).

The physical interpretation of the first optimum compaction parameter \( \hat{T} \) is quite clear. It is the first-order moment of the energy distribution of \( x \) in the time domain: \( \hat{T} = T_1 \).

In order to arrive at an interpretation of \( \hat{\Omega} \), we transform equation (4.7) to the Fourier domain which gives

\[
\hat{\Omega} = \frac{\int_{-\infty}^{\infty} \omega |X(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |X(\omega)|^2 d\omega}. \quad (4.9)
\]

Thus, we see that \( \hat{\Omega} \) equals the first-order moment of the energy distribution of \( x \) in the Fourier domain: \( \hat{\Omega} = \Omega_1 \).

From (6) and (8) we infer that for real-valued signals the optimum modulation \( \hat{\Omega} \) equals zero. This simplifies the results [21] but at the same time tells us that the proposed procedure is not quite adequate for strong-oscillating real-valued signals. In these cases one can better apply the proposed procedure to the associated analytic signal.

Finally, we transform equation (4.8) partially to the Fourier domain, which gives

\[
\hat{\sigma}^4 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{(\omega - \hat{\Omega})^2 |X(\omega)|^2 d\omega}{\langle (t - \hat{T})x, (t - \hat{T})x \rangle}. \quad (4.10)
\]

Since \( \hat{T} = T_1 \) and \( \hat{\Omega} = \Omega_1 \), the squared optimum scale \( \hat{\sigma}^2 \) is the geometric mean of the second-order central moment of the energy distribution in the frequency domain.
and the inverse of the second-order central moment of the energy distribution in the
time domain.

We will now consider the minimum compaction measure. We substitute the optimum
compaction parameters and obtain

\[
Q(x; \hat{\theta}) = \sqrt{\left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} (\omega - \hat{\Omega})^2 |X(\omega)|^2 d\omega \right\} \left\{ \int_{-\infty}^{\infty} (t - \hat{T})^2 |x(t)|^2 dt \right\} - \frac{1}{2} \langle x, x \rangle},
\]

where \( \hat{\theta} \) denotes the optimum compaction parameters \( \hat{\sigma}, \hat{T}, \) and \( \hat{\Omega}. \) We see that the
minimum center of the energy distribution in the transform domain, that is,

\[
\zeta(x; \hat{\theta}) = \min_{\theta}\{\zeta(x; \theta)\} = Q(x; \hat{\theta})/\langle x, x \rangle,
\]

equals the time-bandwidth product minus one half:

\[
\zeta[\hat{\theta}] = TBWP - 1/2.
\]

Again, we note the nice symmetry in the time and frequency domain in results
(4.6), (4.9), (4.10), and (4.11).

\section*{4.6 Discussion}

We have considered functions having finite energy and finite effective support ex-
panded in a modulated Hermite series. In order to obtain a compact series expansion,
we introduced a compaction criterion. We used the free parameters in the modulated
Hermite series to minimize this criterion.

The chosen compaction criterion equals the first-order moment of the energy dis-
tribution in the transform domain. This specific criterion, and its minimization, is
especially suitable for the considered series expansion since it allows us to make use
of the differential equation associated with the modulated Hermite functions. As a
consequence, simple expressions for the optimum compaction parameters evolve.

The resulting optimum compaction parameters have simple physical interpreta-
tions. Furthermore, we have shown that the minimum compaction criterion that is
obtained using these parameters is directly related to the time-bandwidth product.
We note that, thanks to the introduction of the modulation, the optimum parameters
obtain the above-mentioned physical interpretations; see [20].

The considered method of trimming the parameters to obtain compaction can be used in,
for example, coding applications and pattern recognition. In the latter case, similarity
between patterns can be established on the basis of the transform
domain parameters regardless of the scale, position, or modulation of the underlying functions.

Since the minimum compaction measure for the modulated Hermite series expansion is related to the time-bandwidth product, we have a link to the work of Gabor. We have shown that the center of the energy in the modulated Hermite transform has a minimum $\zeta[\hat{\theta}]$ which is approximately equal to the time-bandwidth product. This means that with a number of $\zeta[\hat{\theta}]$ parameters, we have captured roughly 50\% of the energy of the original function and, as a rule of thumb, we expect that with $2\zeta[\hat{\theta}]$ parameters we can obtain an accurate description of the original signal. This is close to the lower bound on the number of relevant parameters as hypothesized by Gabor.

Therefore, we will compare the Gabor expansion with a modulated Hermite series expansion. Consider the class of functions having the same time-bandwidth product $N$. Then, Gabor states, we would need approximately $N$ parameters to code the signal in a Gabor expansion. We note, however, that the Gabor expansion is non-orthogonal and that it is a numerically difficult task to calculate the expansion coefficients [4]. Furthermore, it is not clear how the time-frequency grid associated with the Gabor expansion must be chosen; nor do we know how close to (or how far from) the original an $N$th term expansion may be.

On the other hand, we have the modulated Hermite series expansion. This is an orthonormal series expansion, and the coefficients can be easily calculated. The optimum free parameters have been established in this chapter, and an upper bound for the quadratic truncation error can be given.

In conclusion, since the modulated Hermite series expansion is more amenable to analysis and it is expected to have better numerical properties, we suggest that this expansion can offer an interesting alternative to the Gabor expansion, assuming a finite TBWP. However, since the modulated Hermite functions can be centered at one specific location in the time and frequency domain only, they are less suited for the approximation of continuously emitted signals as well as for signals having various clear-distinguishable resonant frequencies.
Chapter 5

Design of causal systems by Kautz transformations

The purpose of this chapter is to provide signal analysis by causal orthonormal systems, based on simple rational transfer functions.

The outline of this chapter is as follows. In Section 5.1 we introduce the modulated Laguerre expansions. In Laplace domain they are Laguerre functions that allow complex poles. The modulated Laguerre expansions can be used in approximating functions with clearly resonant character. In this section we derive properties of the expansions and compute the optimal parameters according to the enforced convergence criterion.

Having the optimal pole $p = -\sigma/2 + j\Omega$ for the modulated Laguerre expansions, the Laguerre series describes compactly a function that is to be approximated, if it has a single side-band modulated low-pass nature. However, on qualitative grounds one can expect that Kautz series of more general type achieves an even more compact representation in the case of approximating a bandpass filter.

In Section 5.2 we consider the Kautz expansions. The modulated Laguerre system restricted to $\alpha = 0$ is special case of the Kautz system. First, we show the similarity between the continuous-time and discrete-time. Further in this section, we concentrate mainly on discrete-time. A general Kautz system does not adhere a fixed difference/differential equation. So the optimization problem has to be treated in a different way. In this section we propose two ways to optimize the parameters.

For functions with clearly resonant character, we restrict to a Kautz series having a repeated complex-conjugated pole pair. From the function $x$ we construct the function $y$ by the Hilbert transformation, thus $x = (y + y^*)/2$. Next, we use this complex function $y$ to find the optimal parameter in a complex Laguerre series. Subsequently, we take this pole, and its complex conjugate, as an estimate for the optimal parameters in the Kautz series.

Another approach to treat the general Kautz systems is considered by the recursive enforced convergence criterion. We determine first the optimal parameter $\hat{p}$ for $x$ in the Laguerre system. Next, we consider the Kautz system where $p_k = \rho$, for
Design of causal systems by Kautz transformations

$k \geq 1$ and $p_0 = \hat{p}$. This system adheres a difference/differential equation, where the optimal $\rho$ can be determined according to the enforced convergence criterion. Subsequently, we put $p_0 = \hat{p}, p_1 = \hat{p}, p_k = \rho, k \geq 2$, and so on we determine the next $p_k$. Recursively, we have a suboptimal solution for the Kautz system. Of course, combinations of the first and second method can be taken. In Section 5.3 we conclude the Chapter with a discussion.

5.1 Optimal parameters in modulated Laguerre expansions

The harmonically modulated Laguerre functions constitute an orthonormal basis in the Hilbert space of square-integrable functions $\mathcal{H} = L^2_\sigma(0, \infty)$. This basis comprises three free parameters, namely a modulation, a scale and an order factor.

We will recall the enforced convergence criterion (see Subsection 5.1.1 and [20,87]), and will show that this specific measure results in simple, explicit expressions for the free parameters in a modulated Laguerre series expansion as a function of the signal which is to be expanded. This result extends earlier work in this area; see [20] and references therein.

The outline of this section is as follows. First we will present the harmonically modulated Laguerre functions and look at some of their properties. Second we will consider a Laguerre series expansion of an arbitrary function in the admissible space $\mathcal{D}_c$ and recall the enforced convergence criterion. Third, we will calculate the optimum free parameters (modulation, scale, and order). Finally, we conclude the section by an academic example.

5.1.1 Definitions and properties

We start with the definition and some properties of the modulated Laguerre functions.

Definition 5.1.1 The generalized Laguerre polynomials, the generalized Laguerre functions and the modulated Laguerre functions are defined by

a) the generalized Laguerre polynomials $l_n^\alpha$:

$$l_n^\alpha(t) = \sum_{m=0}^{n} (-1)^m \binom{n + \alpha}{n - m} \frac{1}{m!} t^m,$$

b) the generalized Laguerre functions $L_n^\alpha$:

$$L_n^\alpha(t; \sigma) = \sqrt{\sigma} \sqrt{w(\sigma t)} l_n^\alpha(\sigma t) / h_n,$$
5.1 Optimal parameters in modulated Laguerre expansions

The modulated Laguerre functions $\phi_n[\theta]$:

$$\phi_n(t; \theta) = e^{\sigma t} L_n^\alpha(t; \sigma),$$

with $\theta = (\alpha, \sigma, \Omega) \in \Theta = \{ (\alpha, \sigma, \Omega) | \Omega \in \mathbb{R}, \alpha > -1, \sigma > 0 \}$, and $(\alpha, \sigma, \Omega)$ being the modulation, order and scale parameter, respectively. Further the associated weight function $w$ and the normalization constant [20] are:

$$w(t) = t^\alpha e^{-t}$$

$$h_n = \sqrt{\frac{\Gamma(\alpha + n + 1)}{n!}}.$$

The modulated Laguerre functions can be easily generated using a recurrence relation:

$$\phi_0(t; \theta) = \sqrt{\frac{\Gamma(\alpha + 1)}{\Gamma(\alpha + 1)}} (\sigma t)^{\alpha/2} e^{-\sigma t/2} e^{\sigma t}$$

$$\phi_1(t; \theta) = \frac{-\sigma t + \alpha + 1}{\sqrt{\alpha + 1}} \phi_0(t; \theta),$$

and for $n \geq 2$

$$\phi_n(t; \theta) = \frac{-\sigma t + \alpha + 2n - 1}{\sqrt{(\alpha + n)n}} \phi_{n-1}(t; \theta) - \sqrt{\frac{(\alpha + n - 1)(n - 1)}{n(\alpha + n)}} \phi_{n-2}(t; \theta),$$

though this is not without numerical problems in the case of large $n$. For the modulated Laguerre functions we can explicitly write

$$\phi_n(t; \theta) = \sum_{m=0}^{n} a_{n,m}^\alpha (\sigma t)^{m+\alpha/2} e^{\sigma t},$$

with

$$a_{n,m}^\alpha = \sqrt{\frac{n!}{\Gamma(n + \alpha + 1)}} (-1)^m \left( \frac{n + \alpha}{n - m} \right) \frac{1}{m!}.$$

for $m \leq n$ and $a_{n,m}^\alpha = 0$ for $m > n$; and:

$$p = -\frac{1}{2} \sigma + \mathfrak{A}.\Omega.$$
In Laplace domain:

\[
\phi_n(s; \theta) = \mathcal{L}\{\phi_n[\theta]\}(s) = \frac{1}{\sqrt{\sigma}} \sum_{m=0}^{n} \Gamma(m + 1 + \alpha/2) a_{n,m} \left( \frac{\sigma}{s-p} \right)^{m+1+\alpha/2},
\]

where the Laplace transformation is denoted by \(\mathcal{L}\) and defined in \(L_2(\mathbb{R}^+)\) as:

\[
X(s) = \mathcal{L}\{x\}(s) = \int_{0}^{\infty} x(t) e^{-st} dt.
\]

We note that the Laplace transforms of the modulated Laguerre functions can be expressed in terms of a hypergeometric functions \([62]\):

\[
\phi_n(s; \theta) = \frac{1}{\sqrt{\sigma}} \sqrt{\frac{\Gamma(\alpha+n+1)\Gamma(\alpha/2+1)}{\Gamma(n+1)}} \left( \frac{\sigma}{s-p} \right)^{\alpha/2 + 1} \times
2F_1(-n, \alpha/2 + 1; \alpha + 1; \frac{\sigma}{s-p}),
\]

where \(pF_q\) is a hypergeometric series defined by:

\[
pF_q(\alpha_1, \alpha_2, \ldots, \alpha_p; \beta_1, \beta_2, \ldots, \beta_q; z) := \sum_{n=0}^{\infty} \frac{(\alpha_1)_n(\alpha_2)_n \cdots (\alpha_p)_n}{(\beta_1)_n(\beta_2)_n \cdots (\beta_q)_n} \frac{z^n}{n!}
\]

and

\[
(\alpha)_n := \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)}.
\]

The parameters \(\sigma\) and \(\Omega\) are related to the real and the imaginary parts, respectively, of the pole of the Laplace transform \(\phi_n\). The parameter \(\alpha\) is related to the multiplicity of the pole \(p\) in \(\phi_n\).

**Lemma 5.1.2** The system \(\{\phi_n[\theta]|n \in \mathbb{N}\}\) of modulated Laguerre functions constitutes an orthonormal basis in \(H = L_2(0, \infty)\).

We note that Lemma 5.1.2 is not a statement about the rate of convergence of such series. This is, however, a major issue in practical applications since it determines the compaction of the expansion. There are many properties related to the modulated Laguerre functions. These essentially follow from all the properties associated with the Laguerre polynomials. We will consider the differential equation only, since this is directly related to the enforced convergence criterion that will be introduced in the next section.
5.1 Optimal parameters in modulated Laguerre expansions

Lemma 5.1.3 The modulated Laguerre functions \( \phi_n[\theta] \) adhere to the following ordinary differential equation on \( \mathbb{R}^+ \)

\[ A_2(t; \theta) \frac{d^2 \phi_n[\theta]}{dt^2} + A_1(t; \theta) \frac{d \phi_n[\theta]}{dt} + A_0(t; \theta) \phi_n[\theta] = n \phi_n[\theta], \tag{5.1} \]

where

\[
\begin{align*}
A_0(t; \theta) &= \frac{(\alpha - \sigma t)^2}{4\sigma t} - \sigma + 2j\Omega + 2\Omega^2 t, \\
A_1(t; \theta) &= -1 + \frac{2j\Omega t}{\sigma}, \\
A_2(t; \theta) &= -\frac{t}{\sigma}.
\end{align*}
\]

The separation coefficients, see Chapter 3, for \( A_k(t; \theta) \) are given by:

\[
\begin{align*}
B_0(\theta) &= \left( \frac{\alpha^2}{4\sigma}, \frac{\sigma^2 + 4\Omega^2}{4\sigma}, -\frac{\alpha + 1}{2} + \frac{\Omega}{\sigma} \right), \\
C_0(t) &= \left( \frac{1}{t}, t, 1 \right), \\
B_1(\theta) &= \left( \frac{2j\Omega}{\sigma}, -\frac{1}{\sigma} \right), \\
C_1(t) &= (t, 1), \\
B_2(\theta) &= -\frac{1}{\sigma}, \\
C_2(t) &= t,
\end{align*}
\]

where \( J_0 = 3, J_1 = 2, \) and \( J_2 = 1. \) We recall the Definition 3.1.1 of the admissible space \( \mathcal{D}_c: \)

\[ x \in \mathcal{D}_c \iff \sum_{n \in \mathbb{N}} \eta_n[\theta] |\langle x, \phi_n[\theta] \rangle|^2 < +\infty, \]

where, in this case, \( \eta_n[\theta] = n. \) The admissible space \( \mathcal{D}_c \) describes the domain of the quadratic functional \( \mathcal{Q}[\theta]: \)

\[
\mathcal{Q}(x; \theta) = \sum_{k=0}^{2} \sum_{j=1}^{J_k} \left( C_{k,j}(t) \left( \frac{d}{dt} \right)^k x, x \right) B_{k,j}(\theta) \\
= \frac{1}{\sigma} \langle tx', x' \rangle + \frac{2j\Omega}{\sigma} \mathcal{G} \{ \left( tx, x' \right) \} + \left( \frac{\Omega^2}{\sigma} + \frac{\Omega}{2} \right) \langle tx, x \rangle - \\
\left( \frac{1}{\sigma} + \frac{\Omega}{2} \right) \langle x, x \rangle + \frac{\alpha^2}{4\sigma} \langle \frac{1}{t} x, x \rangle,
\]

\[ \Box \]
where $\Im$ denotes the imaginary part. We define the admissible space $\mathcal{D}_c \subseteq \mathcal{H}$ for the modulated Laguerre system in $\mathcal{H} = L_2(0, \infty)$ as:

$$\mathcal{D}_c = \left\{ x | (\sqrt{t} + \frac{1}{\sqrt{t}})x, (1 + \sqrt{t}) \frac{dx}{dt} \in L_2(0, \infty) \right\}.$$ 

These functions are characterized by finite energy and finite center of the energy.

### 5.1.2 Criterion and results

Since the Laguerre functions form an orthonormal basis in $L_2(0, \infty)$, every function in the admissible space $\mathcal{D}_c$ can be expressed in this basis. The transformation $\mathcal{Y}[\theta] : x \mapsto \langle x, \phi_n[\theta] \rangle$ is called the modulated Laguerre transformation, for $x \in L_2(0, \infty)$. In the literature $\mathcal{Y}[\theta] x$ is called the expansion coefficients, the Laguerre spectrum, or the Laguerre transform. We refer to it as the Laguerre transform. The relation between $Q[\theta]$ and $\mathcal{Y}[\theta]$ is given by

$$Q(x; \theta) = \sum_{n \in \mathbb{N}} | \langle \mathcal{Y}[\theta] x \rangle (n) |^2,$$

for $x \in \mathcal{D}_c$, which is the compaction measure for modulated Laguerre expansions. The physical interpretation of this measure is obvious: it is the center of the energy distribution in the transform domain. The Laguerre transformation $\mathcal{Y}[\theta]$ has three free variables ($\alpha$, $\sigma$, and $\Omega$) which are used to make the transformation $\mathcal{Y}$ as compact as possible.

**Minimization Problem 5.1.4** Let $x \in \mathcal{D}_c$, minimize

$$Q(x; \theta) = \frac{1}{\sigma} \langle tx', x' \rangle + 2\Omega \Im \{ \langle tx, x' \rangle \} + (\frac{1}{\sigma^2} + \frac{\Omega^2}{\Omega}) \langle tx, x \rangle -$$

$$(\frac{1}{2} + \frac{\Omega}{2}) \langle x, x \rangle + \frac{\Omega^2}{\sigma^2} \langle i^2 x, x \rangle,$$

for a given $x \in \mathcal{D}_c$.

We note that $\lim_{\theta \to 0} Q(x; \theta) = +\infty$, for $x \in \mathcal{D}_c$. So the candidates for the minimum are obtained by:

$$\nabla_{\theta} Q(x, \theta) = \left( \frac{\partial Q}{\partial \alpha}(x; \theta), \frac{\partial Q}{\partial \sigma}(x; \theta), \frac{\partial Q}{\partial \Omega}(x; \theta) \right) = 0,$$

for a given $x \in \mathcal{D}_c^*$. 

**Definition 5.1.5** We define the optimum compaction parameters $\hat{\theta} = (\hat{\alpha}, \hat{\sigma}, \hat{\Omega})$ for a given function $x \in \mathcal{D}_c$ as those $\alpha$, $\sigma$, and $\Omega$ yielding the minimum compaction measure.
5.1 Optimal parameters in modulated Laguerre expansions

**Theorem 5.1.6** For $x \in \mathcal{D}_c$ the optimum compaction parameters are given by

$$
\hat{\Omega} = -\frac{\Im\langle tx, x' \rangle}{\langle tx, x \rangle} \quad (5.3)
$$

$$
\hat{\sigma}^2 = 4 \frac{\langle \frac{1}{t} x, x \rangle \bigl(\langle tx, x \rangle \langle tx', x' \rangle - (\Im\langle tx, x' \rangle)^2 \bigr)}{\langle tx, x \rangle \bigl(\langle tx, x \rangle \langle \frac{1}{t} x, x \rangle - \langle x, x \rangle^2 \bigr)} \quad (5.4)
$$

$$
\hat{\alpha} = \hat{\sigma} \langle x, x \rangle / \langle \frac{1}{t} x, x \rangle. \quad (5.5)
$$

**Proof:** Let $x \in \mathcal{D}_c^*$. Differentiation of (5.2) with respect to $\Omega, \alpha$, and $\sigma$ and equating this to zero leads to the above results. Considering the second-order derivatives, it follows that

$$
\frac{\partial^2 Q}{\partial (\alpha, \sigma, \Omega)^2} (x; \hat{\theta}) = 2 \hat{\sigma} \langle \frac{1}{t} x, x \rangle \left\{ \langle tx, x \rangle \langle tx', x' \rangle - \Im\langle tx, x' \rangle^2 \right\} \geq 0.
$$

So these parameters yield the minimum of $Q(x; \hat{\theta})$.

From (5.3) we infer that for real-valued signals the optimum modulation $\hat{\Omega}$ equals zero. This simplifies the results [21] but at the same time tells us that the proposed procedure is not quite adequate for strong-oscillating real-valued signals. In these cases one can better apply the proposed procedure to the associated analytic signal.

We will now consider the minimum compaction measure. We substitute the optimum compaction parameters and obtain

$$
Q(x; \hat{\theta}) = \frac{2}{\hat{\sigma}} \left\{ \frac{\langle tx', x' \rangle - \hat{\Omega}^2 \langle tx, x \rangle}{\langle x, x \rangle} \right\} - \frac{1}{2},
$$

where $\hat{\alpha}, \hat{\sigma}$, and $\hat{\Omega}$ denote the optimum compaction parameters.

**Example 5.1.7** We consider the following example:

$$
x(t) = \nu e^{-\lambda t},
$$

where $\nu \in \mathbb{R}, \nu > 0$ and $\lambda \in \mathbb{C}, \Re(\lambda) > 0$. To give the optimal parameters for $x$, according to the enforced convergence criterion, the following integral calculations
are needed:

\[
\langle x, x \rangle = \frac{\Gamma(2\nu + 1)}{(\lambda + \lambda^*)^{2\nu + 1}}, \\
\langle tx, x \rangle = \frac{\Gamma(2\nu + 2)}{(\lambda + \lambda^*)^{2\nu + 2}}, \\
\langle \frac{1}{t} x, x \rangle = \frac{\Gamma(2\nu)}{(\lambda + \lambda^*)^{2\nu}}, \\
\langle tx', x' \rangle = \frac{\nu(\lambda - \lambda^*) - \lambda^*}{(\lambda + \lambda^*)^{2\nu + 2}} \Gamma(2\nu + 1), \\
\Re\langle tx, x' \rangle = -\frac{j(\lambda - \lambda^*)/2}{(\lambda + \lambda^*)^{2\nu + 2}} \Gamma(2\nu + 2), \\
\langle tx', x' \rangle = -\frac{\nu^2(\lambda + \lambda^*)^2 + 2\nu(2\nu + 1)\lambda\lambda^*}{(\lambda + \lambda^*)^{2\nu + 2}} \Gamma(2\nu).
\]

The optimal parameters, according to Theorem 5.1.6, are:

\[
\hat{\Omega} = \frac{\lambda - \lambda^*}{2} = -\Im(\lambda) \\
\hat{\sigma} = \lambda + \lambda^* = 2\Re(\lambda) \\
\hat{\alpha} = 2\nu.
\]

This is exactly what we expected where the expansion of \( x \) is described by the first term of the modulated Laguerre series:

\[
x(t) = \langle x, \phi_0[\tilde{\theta}] \rangle \phi_0(t; \tilde{\theta}).
\]

The minimum enforced convergence criterion in this case is zero:

\[
Q(x; \tilde{\theta}) = 0.
\]

\[\Box\]

### 5.2 Optimal parameters in Kautz series expansions

For a function with a clear resonant character, we first apply the enforced convergence criterion to a Laguerre series described by a complex pole. In practice, this is not a very interesting case: only for complex signals we will find an (optimal) complex pole; in the (usual) case of a real signal or system that is to be approximated the procedure gives a real-valued parameter.
However, the results can be applied as follows. Suppose we have a function $x$ having a pronounced resonant character. As an approximation of $x$ we use a truncated Kautz series. We restrict our attention to a Kautz series having a repeated complex-conjugated pole pair. From the function $x$ we construct the function $y$ by the Hilbert transformation, thus $x = (y + y^*)/2$. Next, we use this complex function $y$ to find the optimal parameter in a complex Laguerre series. Subsequently, we take this pole (and its complex conjugate) as an estimate for the optimal parameters in the Kautz series.

By an example, we show that such procedure can give results close to the optimal parameterization defined by a quadratic error criterion. If, however, the results of the proposed procedure are not satisfactory, the obtained estimated pole can always be used as the initial value in some refined optimization procedure.

### 5.2.1 Kautz and Laguerre series

We start with the definition of discrete-time Kautz systems.

**Definition 5.2.1** Let $\{p_n\}_{n \in \mathbb{N}_0}$ be a sequence in $\mathbb{D}$ with the $p_n$’s not necessarily distinct. Define the natural number $j_n$ by $j_n = \sum_{t=0}^{n-1} \delta_{p_n, p_t}$

$$
\delta_{p_n, p_t} = \begin{cases} 
1 & \text{if } p_n = p_t, \\
0 & \text{otherwise.}
\end{cases}
$$

Let $h_n[\theta] \in l_2(\mathbb{N}_0)$ be defined as follows $h_n(t; \theta) = (t^{j_n})p_n^{t-j_n}$, $t, n \in \mathbb{N}_0$. Recall that $\binom{j_n}{t} = 0$ if $t > j_n$. The system $\{h_n[\theta]| n \in \mathbb{N}_0\}$ is linearly independent. Let $\{\phi_n[\theta]| n \in \mathbb{N}_0\}$ be the result of the Gram-Schmidt orthonormalization procedure applied to $\{h_n[\theta]| n \in \mathbb{N}_0\}$. The obtained system $\{\phi_n[\theta]| n \in \mathbb{N}_0\}$ is called a discrete Kautz system.

The system $\{\phi_n[\theta]| n \in \mathbb{N}_0\}$ is a complete orthonormal system in $\ell_2(\mathbb{N}_0)$ if and only if $\sum_n (1-|p_n|)$ diverges [98]. It can be proved by induction that the $z$-transform $\Phi_n[\theta]$ of $\phi_n[\theta]$ is given by [76]:

$$
\Phi_n(z; \theta) = \kappa_n \sqrt{1 - p_n p_n^*} \frac{z}{z - p_n} \prod_{l=0}^{n-1} \frac{1 - z p_l^*}{z - p_l} 
$$

(5.6)

with $\theta = (p_0, p_1, p_2, \ldots)$, $|p_n| < 1$ and $|\kappa_n| = 1$. If all poles are identical, $p_n = \rho$, and real-valued, we have the Laguerre basis [25, 57], in this case we use $\theta = \rho 1$:

$$
\Phi_n(z; \theta) = \sqrt{1 - \rho^2} \frac{z}{z - \rho} \left(1 - \frac{z \rho}{z - \rho}\right)^n.
$$
The functions $\phi_n[\theta]$ are now real-valued. If all poles are identical $p_n = \rho$ with $\Im\{\rho\} \neq 0$, we will call the corresponding orthonormal system the complex Laguerre basis:

$$\Phi_n(z; \theta) = \sqrt{1 - \rho^2} \frac{z}{z - \rho} \left( \frac{1 - z \rho^*}{z - \rho} \right)^n.$$  

In the case that the poles occur as a repeated complex-conjugated pair, that is,

$$p_{2n+1}^* = p_{2n} = p,$$

for $n \in \mathbb{N}$, we will call it a Kautz basis. The more general basis (5.6) is also referred to as a Kautz basis. Since, in this section, we will consider the case of a repeated complex-conjugated pole pair only, there will be no confusion in terminology. In the case of a real function and a Kautz series expansion, we can redefine the Kautz basis in order to obtain a real spectrum. The most common way is to define the functions $\psi_{i,n}(t; \theta)$ with $i = 1, 2$ having $\hat{z}$-transforms

$$\Psi_{i,n}(z; \theta) = K_i \frac{z(z - q_i)}{(z - p)(z - p^*)} \left( \frac{(1 - zp)(1 - zp^*)}{(z - p^*)(z - p^*)} \right)^n\quad \text{(5.7)}$$

with

$$K_1 = \sqrt{(1 + p)(1 + p^*)(1 - pp^*)}/2,$$

$$K_2 = \sqrt{(1 - p)(1 - p^*)(1 - pp^*)}/2,$$

$$q_1 = -1,$$

$$q_2 = 1,$$

$$\theta = (p, p^*, p, p^*, \ldots).$$

The continuous-time Kautz systems are defined by:

**Definition 5.2.2.** Let $\theta = (p_n)_{n \in \mathbb{N}_0}$ be a sequence in $\mathbb{C}^\infty$ with the $p_n$’s not necessarily mutually distinct. Define the natural number $j_n$ as defined in Definition 5.2.1. Let $h_n[\theta] \in L_2(0, +\infty)$ be defined as follows $h_n(t; \theta) = \frac{\phi_n[\theta]}{j_n e^{\theta n t}}, t > 0, n \in \mathbb{N}_0$. The system $\{h_n[\theta]|n \in \mathbb{N}_0\}$ is linearly independent. Let $\{\phi_n[\theta]|n \in \mathbb{N}_0\}$ be the result of the Gram-Schmidt orthonormalization procedure applied to $\{h_n[\theta]|n \in \mathbb{N}_0\}$. The obtained system $\{\phi_n[\theta]|n \in \mathbb{N}_0\}$ is called a continuous Kautz system.
The system \( \{ \phi_n[\theta] \mid n \in \mathbb{N}_0 \} \) is a complete orthonormal system in \( \ell_2(\mathbb{N}_0) \) if and only if \( \sum_n \frac{-\Re(p_n)}{1+|p_n|^2} \) diverges [98]. It can be proved by induction that the Laplace transform \( \Phi_n[\theta] \) of \( \phi_n[\theta] \) is given by [76]:

\[
\Phi_n(s; \theta) = \kappa_n \frac{\sqrt{2\Re(-p_n)}}{s - p_n} \prod_{l=0}^{n-1} \frac{s + p_l^*}{s - p_l},
\]

with \( \theta = (p_0, p_1, p_2, \ldots), p_n \in \mathbb{C}^- \), \( |\kappa_n| = 1 \).

In the following we show the similarity between the discrete-time in \( \ell_2(\mathbb{N}_0) \) and continuous-time in \( L_2(0, \infty) \). Consider the following conformal mapping (Möbius transformation):

\[
s = a \frac{z - 1}{z + 1},
\]

where the parameter \( a \) is an arbitrary positive real number. This conformal mapping transforms the unit circle \( U \) in the \( z \)-plane onto the imaginary axis, in the \( s \)-plane, the disc \( D \) into the left half plane \( \mathbb{C}^- = \{ s \mid \Re(s) < 0 \} \), and therefore \( \mathbb{E} \) (the exterior of \( U \cup D \)) into the right half plane \( \mathbb{C}^+ = \{ s \mid \Re(s) > 0 \} \). The inverse of this conformal mapping is given by

\[
z = \frac{a + s}{a - s}.
\]

Note that a contour integral over the unit circle \( U \) in the \( z \)-plane is transformed into a line integral over the imaginary axis in the \( s \)-plane. In particular,

\[
\langle \Phi, \Psi \rangle = \frac{1}{2\pi i} \int_U \Phi(z) \Psi^*(1/z^*) \frac{dz}{z} = \frac{1}{2\pi i} \int_{-\infty}^{+\infty} \Phi \left( \frac{a + s}{a - s} \right) \Psi^* \left( \frac{a - s^*}{a + s^*} \right) \frac{2a}{(a - s)(a + s)} ds.
\]

This leads naturally to the definition of the transformation \( B \) on the Hardy space \( \mathcal{H}_2(\mathbb{E}) \)

\[
(B\Phi)(s) = \frac{\sqrt{2a}}{a + s} \Phi \left( \frac{a + s}{a - s} \right),
\]

for \( \Phi \in \mathcal{H}_2(\mathbb{E}) \). So the transformation \( B \) maps the Hardy space \( \mathcal{H}_2(\mathbb{E}) \) into the Hardy space \( \mathcal{H}_2(\mathbb{C}^+) \) of analytic functions in \( \mathbb{C}^+ \) that vanish at infinity and that are square
Lebesgue integrable on \( \mathbb{R} \). The Hardy space \( \mathcal{H}_2(\mathbb{C}^+) \) is a Hilbert space with inner product

\[
< \Phi, \Psi > = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \Phi(s) \Psi^*(s^*) ds
\]

for \( \Phi, \Psi \in \mathcal{H}_2(\mathbb{C}^+) \). From (5.12) and (5.13) we have

\[
< \Phi, \Psi > = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (B\Phi)(s) (B\Psi)^*(s^*) ds
\]

\[
= < B\Phi, B\Psi > .
\]  

Thus \( B \) defines an isometry from \( \mathcal{H}_2(\mathbb{R}) \) into \( \mathcal{H}_2(\mathbb{C}^+) \). Further, we consider the operator \( \bar{B} \) from \( \mathcal{H}_2(\mathbb{C}^+) \) into \( \mathcal{H}_2(\mathbb{R}) \), defined by

\[
(\bar{B}\Psi)(z) = \frac{\sqrt{2\alpha}}{z + 1} \Psi \left( a \frac{z - 1}{z + 1} \right),
\]

for \( \Psi \in \mathcal{H}_2(\mathbb{C}^+) \). For each \( \Psi \in \mathcal{H}_2(\mathbb{C}^+) \) it follows

\[
B(\bar{B}\Psi) = \Psi.
\]

Thus the operator \( B \) is unitary from \( \mathcal{H}_2(\mathbb{R}) \) onto \( \mathcal{H}_2(\mathbb{C}^+) \) with inverse \( B^{-1} = \bar{B} \).

Using the unitarity of the Laplace transformation \( \mathcal{L} : L_2(\infty) \to \mathcal{H}_2(\mathbb{C}^{-}) \), and the \( z \)-transformation \( Z : \ell_2(\mathbb{N}_0) \to \mathcal{H}_2(\mathbb{R}) \), the operator \( B \) associates discrete-variable functions in \( \ell_2(\mathbb{N}_0) \), via their \( z \)-transforms, uniquely with continuous-variable functions in \( L_2(0, \infty) \), via their Laplace transforms 5.1. The operator \( \mathcal{L}^{-1}BZ \) is a one-to-one transformation from \( \ell_2(\mathbb{N}_0) \) onto \( L_2(0, \infty) \).
5.2 Optimal parameters in Kautz series expansions

5.2.2 Optimal discrete Laguerre parameter

In this subsection, we show that the optimal parameters for the discrete-time Laguerre series expansions can directly be derived from the optimal ones for the modulated continuous-time Laguerre series, where $\alpha = 0$. Therefore, we show first that the transformation $Z^{-1}B^{-1}L$ applied to a modulated Laguerre function $\phi_n^\alpha[\theta]$, $(\alpha = 0)$, leads to the $n$-th element of the discrete Laguerre system. This is also the case for the Kautz system. We recall the differential equation for modulated Laguerre functions ($\alpha = 0$):

$$-\frac{d}{y''[\theta]} + \frac{s}{\sigma}y'[\theta] + \left[\frac{(\sigma^2/4 + \Omega^2)s - \sigma/2 + j\Omega}{\sigma}\right]y[\theta] = ny[\theta].$$

Applying the Laplace transformation to this differential equation we obtain:

$$\frac{1}{\sigma}(s - \sigma/2 - j\Omega) \left[(s + \sigma/2 - j\Omega)Y[\theta] + Y[\theta]\right] = nY[\theta],$$

where $Y$ is the Laplace transform of $y$. Substitute $p = -\sigma/2 + j\Omega$, we have the simplified form:

$$-\frac{1}{p+p^*}(s + p^*) \left[(s - p)Y[\theta]\right]' = nY[\theta].$$

Now we transfer the equation from its description in the Laplace domain to its description in the $z$-domain using $B$, it results in the differential equation:

$$-\frac{1}{2a(p+p^*)}((a+p^*)z - a + p^*) \left[((a - p)z - a - p) G^*[\theta] + (a + p)z^{-1}G\right] = nG[\theta],$$

where

$$G(z) = (B^{-1}Y)(z),$$

$$(B^{-1}Y^*)(z) = \frac{1}{2a}(z + 1) \left((z + 1)G^*(z) - z^{-1}G(z)\right).$$

In the $z$-domain the parameter $p$ has the form $\rho = (a + p)/(a - p)$. Therefore we substitute $p = a(\rho - 1)/(\rho + 1)$, in order to obtain:

$$\frac{1}{1-\rho z} \left[(z - \rho)G^*[\theta] + \rho z^{-1}G[\theta]\right] = nG[\theta]. \quad (5.15)$$

We recognize that (5.15) as the differential equation of the $z$-transformed Laguerre functions. Converting the equation from $\mathcal{H}_2(\mathbb{E})$ into $\ell_2(\mathbb{N}_0)$, finally we get the second-order difference equation for discrete-time Laguerre functions:

$$A_2(t; \theta)\phi_n[\theta] + A_1(t; \theta)\phi_n[\theta] + A_0(t; \theta)\phi_n^{-1}[\theta] = n\phi_n[\theta], \quad (5.16)$$
where

\[ A_0(t; \theta) = -\frac{\rho^k}{1 - \rho^p}, \]
\[ A_1(t; \theta) = \frac{(1 + \rho^p)t + \rho^p}{1 - \rho^p}, \]
\[ A_2(t; \theta) = -\frac{\rho^p(t + 1)}{1 - \rho^p}, \]

and \( \theta = \rho \), the free parameter.

Taking \( \alpha = 0 \) in (5.2), and minimizing \( Q(x_c; \theta) \), \( x_c \in D_c^* \), with respect to \( \Omega \) and \( \sigma \), we obtain the optimal parameters for modulated Laguerre series in the case \( \alpha = 0 \):

\[ \hat{\Omega} = -\frac{\Im\langle t x_c, x'_c \rangle}{\langle t x_c, x_c \rangle} \]
\[ \hat{\sigma}^2 = 4 \left( \frac{\Im\langle t x_c, x'_c \rangle}{\langle t x_c, x_c \rangle} \right)^2 - 4 \left( \frac{\Im\langle t x_c, x'_c \rangle}{\langle t x_c, x_c \rangle} \right)^2. \]

for \( x_c \in D_c \). Applying the conformal mapping, (5.10), to these optimal parameters, (5.17) and (5.18), in the continuous case, we obtain the optimal parameters in the discrete case, and vice versa (using (5.9)).

Let \( x_d \in \ell_2(\mathbb{N}_0) \), we have shown that the discrete-time function \( x_d \) corresponds with a continuous-time function \( x_c \) in \( L_2(0, \infty) \), by \( x_c = \mathcal{L}^{-1}BZx_d \). Using this notation, \( x_d \in \ell_2(\mathbb{N}_0) \) and its corresponding continuous-time version \( x_c = \mathcal{L}^{-1}BZx_d \), there will be no confusion in terminology. Now we have the following procedure to find the optimal pole of the discrete-time Laguerre series given the \( x_d \). We start with an \( x_d \) in the admissible space for discrete-time Laguerre system \( D_d = \{ x| x, \sqrt{2}x \in \ell_2(\mathbb{N}_0) \} \). It can be proved that the transformation \( \mathcal{L}^{-1}BZ \) maps \( D_d \) unitarily into \( D_c \). We write the optimal parameters for \( x_c = \mathcal{L}^{-1}BZx_d \), and derive the optimal parameters for \( x_d = \mathcal{Z}^{-1}B^{-1}\mathcal{L} \), in the discrete-time Laguerre series, independent of the transformations as function of \( x_d \). We know that the transformation \( \mathcal{L}^{-1}BZ \) maps \( D_d \) unitarily into \( D_c \).

\[ \langle x_c, y_c \rangle = \langle \mathcal{Z}^{-1}B^{-1}\mathcal{L}x_c, \mathcal{Z}^{-1}B^{-1}\mathcal{L}y_c \rangle, \]

for \( x_c, y_c \in L_2(0, \infty) \). Further,

\[ \langle t x_c, x'_c \rangle = \langle \mathcal{L}(t x_c), \mathcal{L}(x'_c) \rangle = \langle -X'_c, sX_c \rangle, \]
\[ \langle t x_c, x_c \rangle = \langle -X'_c, X_c \rangle, \]
\[ \langle t x'_c, x'_c \rangle = \langle -sX'_c - X_c, sX_c \rangle. \]
Applying $B^{-1}$ we obtain
\[
\langle -X'_c, sX_c \rangle = \left\langle -\frac{1}{2a} (z + 1) \left[ (z + 1)G' - z^{-1}G \right], a \frac{z - 1}{z + 1}G \right\rangle
\]
\[
= \frac{1}{2} \left\langle (z - 1) \left[ (z + 1)G' - z^{-1}G \right], G \right\rangle,
\]
\[
\langle -X'_c, X_c \rangle = \left\langle -\frac{z + 1}{2a} \left[ (z + 1)G' - z^{-1}G \right], G \right\rangle,
\]
\[
\langle -sX'_c - X_c, sX_c \rangle = \left\langle -a \frac{z - 1}{z + 1} \frac{z + 1}{2a} \left[ (z + 1)G' - z^{-1}G \right] + G, a \frac{z - 1}{z + 1}G \right\rangle
\]
\[
= \frac{a}{2} \left\langle (z - 1)^2G' + (1 - z^{-1})G, G \right\rangle.
\]
This way the relations in $\ell_2(\mathbb{N}_0)$ occur:
\[
\langle l x_c, x'_c \rangle = \frac{1}{2} \left( \langle (t + 1)T x_d, x_d \rangle^* - \langle (t + 1)T x_d, x_d \rangle - \langle x_d, x_d \rangle \right),
\]
\[
\langle l x_c, x_c \rangle = \frac{1}{2a} \left( \langle (t + 1)T x_d, x_d \rangle^* + \langle (t + 1)T x_d, x_d \rangle + \langle (2t + 1)x_d, x_d \rangle \right),
\]
\[
\langle l x'_c, x'_c \rangle = \frac{a}{2} \left( -\langle (t + 1)T x_d, x_d \rangle^* - \langle (t + 1)T x_d, x_d \rangle + \langle (2t + 1)x_d, x_d \rangle \right).
\]
Define for simplicity in notation the measurement\(^1\):
\[
\beta(x_d) := \frac{\langle (t + 1/2)x_d, x_d \rangle}{\langle (t + 1)T x_d, x_d \rangle}.
\] (5.19)
Substitution in $\hat{\Omega}$ and $\hat{\sigma}$ leads to:
\[
\hat{\Omega} = ja \frac{\beta - \beta^*}{\beta + \beta^* + 2\beta \beta^*},
\]
\[
\hat{\sigma}^2 = 16a^2 \frac{\beta \beta^* (\beta \beta^* - 1)}{(\beta + \beta^* + 2\beta \beta^*)^2}.
\]
It can be proved that for $x_d \neq 0$:
\[
|\langle (t + 1)T x_d, x_d \rangle|^2 < |\langle (t + 1/2)x_d, x_d \rangle|^2,
\]
so $|\beta| > 1$. For the optimal parameter $\hat{\rho} = (a + \hat{\rho})/(a - \hat{\rho})$, where $\hat{\rho} = -\hat{\sigma}/2 + j\hat{\Omega}$, we write:
\[
\hat{\rho} = \frac{a - \hat{\sigma}/2 + j\hat{\Omega}}{a + \hat{\sigma}/2 - j\hat{\Omega}} = \frac{a^2 - a\hat{\sigma} + \hat{\sigma}^2/4 + \hat{\Omega}^2}{a^2 - 2ja\Omega - (\hat{\sigma}^2/4 + \hat{\Omega}^2)}
\]
\[
= \beta^* \left( 1 - \sqrt{1 - \frac{1}{\beta \beta^*}} \right),
\]
\(^1\)In the paper [23] we defined $\beta$ as the complex conjugated of (5.19).
where $|\beta^* \left( 1 - \sqrt{1 - \frac{1}{\beta^*}} \right) | < 1$, because $|\beta| > 1$.

**Theorem 5.2.3** In the discrete Laguerre case, the optimum compaction parameter $\rho$ is given by:

$$\hat{\rho} = \beta^* \left( 1 - \sqrt{1 - \frac{1}{\beta^*}} \right),$$  \hspace{1cm} (5.20)

where for $x \in D_d^*$. $\beta$ is defined in (5.19). The minimum compaction measure is

$$Q(x; \hat{\rho}) = \left| \langle (t + 1)T x, x \rangle \sqrt{\beta^* - 1} - \frac{1}{2} \langle x, x \rangle. \right|$$

\[ \square \]

The admissible space $D_d$ for the discrete-time Laguerre series expansions is given by $D_d = \{ x | x, \sqrt{t} x \in \ell_2(\mathbb{N}_0) \}$. The transformation $L^{-1}BZ$ maps $D_d$ unitarily into $D_c$. Note that if $x_d$ is real, then the optimal pole $\hat{\rho}$ is real as well, no matter how resonant the character of $x$ is. In that case, $Q(x_d; \theta)$ will be large since the resonant character can only be mimicked by many (non-resonant) Laguerre functions.

**Remark 5.2.4** In [23] we treated the problem of optimality directly from the second-order difference equation of discrete-time Laguerre function. We determined the functional $Q(x_d, \theta)$, and minimized it with respect to $\rho$, for a fixed $x \in D_d$ different from zero. We took $\rho = \alpha + \beta$, where $\alpha$ and $\beta$ are real-valued. This way the optimal parameter $\hat{\rho}$ according to the enforced convergence criterion can be achieved. The result is identical to the one in Theorem 5.2.3. \[ \square \]

We now try to give an interpretation of the estimated argument (the angle) of $\hat{\rho}$. We have

$$\hat{\rho} = \beta^* \left( 1 - \sqrt{1 - \frac{1}{\beta^*}} \right),$$

and in view of (5.19) we have

$$\arg \hat{\rho} = \arg \beta^* = - \arg \{(t + 1)T x, x\}.$$

We rewrite the expression $\langle (t + 1)T x, x \rangle$ to the frequency domain which yields

$$\arg \hat{\rho}_c = \arg \left\{ \int_{\mathbb{T}} w(\theta) e^{i\theta} d\theta \right\}$$
5.2 Optimal parameters in Kautz series expansions

with

\[ w(\theta) = -|H(e^{i\theta})| \left[ \frac{\partial}{\partial r} \left\{ \frac{1}{\sqrt{r}} H(re^{i\theta}) \right\} \right]_{r=1}. \]

We infer that \( e^{i\theta} \) is multiplied over the unit circle with the function\(^2\) \( w \) and from this the angle is taken. The function \( w \) is intuitively a good measure. If the function \( G(z) \) has poles \( p_G \) (inside the unit circle) than if \( \theta \simeq \arg p_G \) both \( |H(e^{i\theta})| \) and \( \left[ -\frac{\partial}{\partial r} \left\{ \frac{1}{\sqrt{r}} H(re^{i\theta}) \right\} \right]_{r=1} \) will be large.

5.2.3 Optimal Kautz parameter

The solution of minimization problem of the enforced convergence criterion depends on the existence of a difference equation for the (complex) Laguerre series [20, 87]. This leads to the existence of a linear operator \( \mathcal{A}[\theta] \) such that

\[ \mathcal{A}[\theta] \phi_n[\theta] = \eta_n[\theta] \phi_n[\theta] \]

There are Kautz expansions where an operator of this form for the corresponding orthonormal basis does not exist, even though the structure of the Kautz functions is similar to the one of the Laguerre functions in the sense that

\[ \Psi_i, n[\theta] = \Psi_{i, 0}[\theta] \{ A[\theta] \}^n \]

where \( A(z; \theta) \) is the transfer function of an all-pass filter, see (5.7). Thus, differentiation with respect to \( z \) automatically gives rise to the appearance of the desired multiplicative factor \( n \) needed in the linear operator in order that the same procedure can be applied to the Kautz functions. The differential equation satisfied by the \( n \)-th element of the Kautz basis in the \( z \)-domain is:

\[ \Phi_n'[\theta] + \left[ \frac{p_n z^{-1}}{z - p_n} + \sum_{k=0}^{n-1} \frac{1 - p_k p_k^*}{(z - p_k)(1 - z p_k^*)} \right] \Phi_n[\theta] = 0. \]

For simplicity we introduce

\[ B_n(z; \theta) := \prod_{l=0}^{n} (z - p_l)(1 - z p_l^*). \]

Multiplying the difference equation by \((z - p_n) B_n-1(z; \theta)\) we obtain:

\[ (z - p_n) B_n-1(z; \theta) \Phi_n'[\theta] + z^{-1} D_n(z; \theta) \Phi_n[\theta] = 0. \quad (5.21) \]

\(^2\)The function \( w \) is not necessarily positive for all \( \theta \).
where the expression

\[ D_n(z; \theta) := z B_{n-1}(z; \theta) \left[ p_n z^{-1} + \sum_{k=0}^{n-1} \frac{(1 - p_k p_k^*)(z - p_n)}{(z - p_k)(1 - z p_k^*)} \right] \]

is a polynomial of \( z \), dependent on \( \theta = (p_0, p_1, p_2, \ldots) \). If \( D_n(z; \theta) \) can be written as \( D_n(z; \theta) = (\eta_n[\theta] z^{n_0} + c[\theta]) R_n(z; \theta) \), where \( (z - p_n) B_{n-1}(z; \theta)/R_n(z; \theta) \) is a polynomial and \( c[\theta] \) independent of \( n \), then the enforced criterion can be defined as \( Q(x; \theta) := \sum_{n=0}^{\infty} \eta_n[\theta]|\langle x, \phi_n[\theta] \rangle|^2 \), and the problem of optimality can be solved as for the other classical functions. Although this approach seems rather promising, it turns out that only in academic cases such representation of \( D_n(z; \theta) = (\eta_n[\theta] z^{n_0} + c[\theta]) R_n(z; \theta) \) can be obtained.

An intuitive way to estimate the optimal pole in the Kautz series is to look at the amplitude spectrum, search for the dominant peak (assuming it exists) or the center of energy distribution as a function of the frequency \( \theta \) \((0 \leq \theta \leq \pi\) for a real-valued function). This frequency is then identified with the argument of \( p \). Next, one considers the signal as a modulated low-pass signal and identifies \( \theta \) with the optimal Laguerre parameter for the low-pass signal. A slightly different method is considered below.

### 5.2.4 Repeated complex conjugated pole pair

From a real function \( x \) having a clearly resonant character, we construct the complex function \( y \) as its Hilbert transform\(^3\). Next, the complex function \( y \) is written as a complex Laguerre series and the optimal complex Laguerre parameter for \( y \) is calculated. Since

\[
x = \{y + y^*/2 = \frac{1}{2} \left\{ \sum_{n=0}^{\infty} [a_n \phi_n[\theta] + a_n^* \phi_n^*[\theta]] \right\}
\]

and since the set of functions \( \{\phi_n[\theta], \phi_n^*[\theta]|n = 0, \ldots N\} \) spans the same space as \( \{\psi_{1,n}[\theta], \psi_{2,n}[\theta]|n = 0, \ldots, N\} \) we expect the optimal complex parameter for an approximation of \( g \) to be a good candidate for a compact representation of \( x \) in a Kautz series.

**Algorithm 5.2.5** For \( x \in D_a^* \):

1. \( y = \text{hilbert}(x) \);
2. \( \hat{p} = \beta^* \left(1 - \sqrt{1 - \frac{1}{\beta^2}}\right); \)
   % optimal complex Laguerre parameter for \( y \);

\(^3\)The imaginary part of \( y \) is in general non-causal. Only the causal part of \( y \) is considered.
Example 5.2.6 As an example, we take the function $x$ defined by a rational $z$-transform

$$X(z) = \frac{\prod_{i=1}^{3}(z - q_i)}{\prod_{i=1}^{4}(z - p_i)}$$

with

$$q_1 = q_2^* = 0.96e^{\pi i/5.5},$$
$$q_3 = -1,$$
$$p_1 = p_2^* = 0.97e^{\pi i/5},$$
$$p_3 = p_4^* = 0.975e^{3\pi i/6}.$$ 

In view of the complex conjugated pole pairs of the poles and zeros, $x$ is obviously real-valued. In Fig. 5.2 the amplitude spectrum is given. We observe two closely-spaced resonant peaks. We also note that the function $x$ can never be expressed exactly in a finite Kautz series defined by a repeated complex-conjugated pole pair.

As discussed in the previous section, we consider the causal part of the Hilbert transform of $x$, and from this function calculated the optimal complex pole for a complex Laguerre series. This was found that for $\hat{\rho}_c$ we have $|\hat{\rho}_c| = 0.922$ and $\arg\{\hat{\rho}_c\} = 0.562$ (rad).

Next, we used this pole in a Kautz series and calculated the relative modeled energy as a function of the number of sections $N$. By this we mean the number

\[
\text{Relative Modeled Energy} = \frac{\text{Modeled Energy}}{\text{True Energy}}
\]

Figure 5.2: Amplitude spectrum of the function $x$. 

of complex-conjugated pole pairs. Thus, the order of the approximating transfer function equals $2N$.

In Fig. 5.3, the relative energy loss $E_r$ is shown as a function of $N$. We also calculated this loss for a Kautz series where, for each order $N$, the pole was optimized in such a way that a maximum of energy was contained in the first $N$ terms (Fig. 5.3, dashed line). As is to be expected, the latter is always better (in fact, by definition the best in a quadratic sense). The proposed energy compaction criterion yields results which are close to the optimal ones. This is also shown in Table 5.1.

In Figs. 5.4 and 5.5, we plotted the optimal pole radius and angle according to a maximum energy criterion as a function of the number of sections in the Kautz series (dashed lines) together with the optimal pole according to the energy compaction criterion. We note that the optimal pole defined by the enforced convergence criterion is not quite in line with those obtained by a quadratic criterion.

Even though the poles estimated by the two procedures do not seem to agree, the behaviour in terms of the relative loss is similar. One additional term is needed to obtain roughly the same relative loss in a series expansion using the optimal pole defined by the enforced convergence criterion compared to the optimal pole in a quadratic error criterion.
5.2 Optimal parameters in Kautz series expansions

Figure 5.4: Optimal pole radius as a function of the number of expansion terms. Optimality according to the energy compaction criterion (solid line) and the quadratic error (dashed line).

Figure 5.5: Optimal pole angle as a function of the number of expansion terms. Optimality according to the energy compaction criterion (solid line) and the quadratic error (dashed line).

5.2.5 Recursive solution for optimal Kautz parameters

In the case that we are more interested in approximations using different poles in the Kautz system, the following approach is proposed. Suppose we have determined the first \( l \) parameters, \( l \geq 0 \), and let \( \theta_l(p) = (p_0, p_1, \ldots, p_{l-1}, p, p, \ldots) \). In this subsection we restrict to real values of \( p, p_k \in (-1, 1) \), \( k = 0, 1, \ldots , l - 1 \). We show how to determine the optimal parameter recursively. For \( n \geq l \) we have according
Table 5.1: Optimal pole radius and angle as a function of the number of sections, for a quadratic error criterion, together with the relative modeled energy loss $E_{r2}$. The relative modeled energy loss $E_{r1}$ for the optimal pole according to the energy compaction criterion (see text) is given in the last column.

<table>
<thead>
<tr>
<th>order</th>
<th>radius</th>
<th>angle</th>
<th>$E_{r1}$</th>
<th>$E_{r2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.928</td>
<td>0.578</td>
<td>0.152</td>
<td>0.141</td>
</tr>
<tr>
<td>2</td>
<td>0.965</td>
<td>0.575</td>
<td>0.138</td>
<td>0.069</td>
</tr>
<tr>
<td>3</td>
<td>0.937</td>
<td>0.572</td>
<td>0.046</td>
<td>0.018</td>
</tr>
<tr>
<td>4</td>
<td>0.956</td>
<td>0.573</td>
<td>0.011</td>
<td>0.008</td>
</tr>
<tr>
<td>5</td>
<td>0.939</td>
<td>0.571</td>
<td>0.009</td>
<td>0.002</td>
</tr>
</tbody>
</table>

To (5.21):

$$-\frac{(z-p)(1-p^2)}{1-p^2} \Phi'_n[\theta] + \left[ -\frac{p_{r-1}^2+p^2}{1-p^2} + l - \frac{(z-p)(1-p^2)}{1-p^2} h_l(z) \right] \Phi_n[\theta] = 0, \quad (5.22)$$

where

$$h_l(z) = \sum_{k=0}^{l-1} \frac{1-p_k^2}{(z-p_k)(1-p_k z)}.$$

In terms of the differential operator $A[\theta]$: $A[\theta] \Phi_n[\theta] = n \Phi_n[\theta]$, where

$$A[\theta] \Phi_n[\theta] = -\frac{(z-p)(1-p^2)}{1-p^2} \Phi'_n[\theta] + \left[ -\frac{p_{r-1}^2+p^2}{1-p^2} + l - \frac{(z-p)(1-p^2)}{1-p^2} h_l(z) \right] \Phi_n[\theta].$$

In the time-domain this differential equation corresponds with a finite-order difference equation. Since the $z$-transformation $Z$ maps unitarily $l_2(\mathbb{N}_0)$ onto $H_2(\mathbb{E})$, and since the expressions in terms of the differential equation (5.22) are simpler in the $z$-domain than in the time-domain, we consider the enforced convergence criterion in the $z$-domain. So we come to the following problem: Determine $p$ minimizing $Q(x; \theta_l(p)) = \sum_{n=0}^{\infty} n || \langle X, \Phi_n[\theta_l(p)] \rangle ||^2$. Rewriting the expression for $Q(x; \theta_l)$:

$$Q(x; \theta_l) = \sum_{n=0}^{\infty} n \langle X, \Phi_n[\theta_l] \rangle \langle \Phi_n[\theta_l], X \rangle$$

$$= \sum_{n=0}^{l-1} n \langle X, \Phi_n[\theta_l] \rangle \langle \Phi_n[\theta_l], X \rangle + \sum_{n=l}^{\infty} \langle X, \Phi_n[\theta_l] \rangle \langle A[\theta_l] \Phi_n[\theta_l], X \rangle$$

$$= \sum_{n=0}^{l-1} \langle X, \Phi_n[\theta_l] \rangle \langle \Phi_n[\theta_l] - A[\theta_l] \Phi_n[\theta_l], X \rangle + \langle A[\theta_l] X, X \rangle,$$
5.2 Optimal parameters in Kautz series expansions

we see that the enforced convergence criterion in the Kautz system, where the first $l$ parameters are taken fixed, consists of two terms: the term $\langle \mathcal{A}[\theta_1] X, X \rangle$ corresponds to a finite-order difference operator, but we have to add the term $\sum_{n=0}^{l-1} \langle X, \Phi_n[\theta_l] \rangle$ $\langle n\Phi_n[\theta_l] - \mathcal{A}[\theta_1] \Phi_n[\theta_l], X \rangle$, which corresponds to the fixed choice of the parameters $p_0, p_1, \ldots, p_{l-1}$. This is due to the fact that the first $l$ Kautz basis elements $\Phi_k[\theta_l], k = 0, \ldots, l - 1$ are not eigenfunctions of $\mathcal{A}[\theta_1]$, in general. However, the optimization problem can now be solved explicitly. Indeed, for the difference operator $\mathcal{A}[\theta_1]$, we write the quadratic form in the $z$-domain:

$$
\langle \mathcal{A}[\theta_1] X, Y \rangle = b_{11}(X, Y) \frac{p}{1 - p^2} + b_{12}(X, Y) \frac{1}{1 - p^2} + b_{13}(X, Y),
$$

(5.23)

where

$$
b_{11}(X, Y) = \langle X', (1 + z^{-2})Y \rangle - \langle z^{-1}X, Y \rangle + 
\sum_{k=0}^{l-1} (1 - p_k^2) \left[ \langle \frac{1}{z - p_k} X, \frac{z}{z - p_k} Y \rangle + \langle \frac{z}{z - p_k} X, \frac{1}{z - p_k} Y \rangle \right],
$$

$$
b_{12}(X, Y) = -2(\langle X', z^{-1}Y \rangle + \langle X, Y \rangle - 2 \sum_{k=0}^{l-1} (1 - p_k^2) \langle \frac{1}{z - p_k} X, \frac{1}{z - p_k} Y \rangle),
$$

$$
b_{13}(X, Y) = \langle X', z^{-1}Y \rangle + (l - 1) \langle X, Y \rangle + \sum_{k=0}^{l-1} (1 - p_k^2) \langle \frac{1}{z - p_k} X, \frac{1}{z - p_k} Y \rangle.
$$

Replacing $Y$ by $X$ in (5.23), we have:

$$
\langle \mathcal{A}[\theta_1] X, X \rangle = b_{11}(X, X) \frac{p}{1 - p^2} + b_{12}(X, X) \frac{1}{1 - p^2} + b_{13}(X, X),
$$

moreover,

$$
\langle n\Phi_n[\theta_l] - \mathcal{A}[\theta_1] \Phi_n[\theta_l], X \rangle = -b_{11}(\Phi_n[\theta_l], X) \frac{p}{1 - p^2} - b_{12}(\Phi_n[\theta_l], X) \frac{1}{1 - p^2} + 
\sum_{k=0}^{l-1} (1 - p_k^2) \langle \frac{1}{z - p_k} \Phi_n[\theta_l], \frac{1}{z - p_k} X \rangle - b_{13}(\Phi_n[\theta_l], X).
$$

In this way, the enforced convergence criterion becomes:

$$
Q(x; \theta_l) = c_{11}(X) \frac{p}{1 - p^2} + c_{12}(X) \frac{1}{1 - p^2} + c_{13}(X),
$$

(5.24)
where \(c_1(X), c_{12}(X)\) and \(c_{13}(X)\) are independent of \(p\):

\[
c_1(X) = b_{11}(X, X) - \sum_{n=0}^{l-1} \langle X, \Phi_n[\theta] \rangle b_{11}(\Phi_n[\theta], X),
\]

\[
c_{12}(X) = b_{12}(X, X) - \sum_{n=0}^{l-1} \langle X, \Phi_n[\theta] \rangle b_{12}(\Phi_n[\theta], X),
\]

\[
c_{13}(X) = b_{13}(X, X) + \sum_{n=0}^{l-1} \langle X, \Phi_n[\theta] \rangle [n(\Phi_n[\theta], X) - b_{13}(\Phi_n[\theta], X)].
\]

The expression to be minimized for the enforced convergence criterion in (5.24) can be rewritten as:

\[
Q(x; \theta_t) = \frac{1}{2} (c_{12}(X) + c_{11}(X)) \frac{1}{1 - p} + \frac{1}{2} (c_{12}(X) - c_{11}(X)) \frac{1}{1 + p} + c_{13}(X).
\]

So we have

\[
\lim_{p \uparrow 1} Q(x; \theta_t) = \text{sgn} (c_{12}(X) + c_{11}(X)) \times \infty,
\]

\[
\lim_{p \downarrow 1} Q(x; \theta_t) = \text{sgn} (c_{12}(X) - c_{11}(X)) \times \infty.
\]

Since \(Q(x; \theta_t) > 0\), for all \(x \in D_d\) and \(p \in (-1, 1)\). It implies that

\[
(c_{12}(X) + c_{11}(X)) \geq 0,
\]

\[
(c_{12}(X) - c_{11}(X)) \geq 0,
\]

for all \(x \in D_d^c\). If \(c_{12}(X) + c_{11}(X)\) or \(c_{12}(X) - c_{11}(X)\) is equal to zero, we obtain

\[
Q(x; \theta_t) = A/(1 + p) + c_{13}(X),
\]

which means that the optimal \(p\) lies at the boundaries \(\pm 1\). This means that \(p\) has to be chosen near \(\pm 1\). If both \(c_{12}(X) + c_{11}(X)\) and \(c_{12}(X) - c_{11}(X)\) are equal to zero, then we have \(Q(x; \theta_t) = c_{13}(X)\). So \(Q(x; \theta_t)\) is constant for fixed \(x\). In this case any \(p \in (-1, 1)\) is optimal to minimize \(Q(x; \theta_t)\). In the following, we assume the case \(c_{12}(X) + c_{11}(X)\) and \(c_{12}(X) - c_{11}(X)\) are both positive.

As a consequence of \(\lim_{p \uparrow 1} Q(x; \theta_t) = \lim_{p \downarrow 1} Q(x; \theta_t) = +\infty\), the enforced convergence criterion has a minimum in \((-1, 1)\), which can be obtained by the derivative of \(Q(x; \theta_t)\) with respect to \(p\):

\[
\frac{d}{dp} Q(x; \theta_t) = \frac{c_{11}(X)p^2 + 2c_{12}(X)p + 1}{(1 - p^2)^2}.
\]

The derivative has one unique root in \((-1, 1)\), namely

\[
\hat{p} = \frac{c_{12}(X)}{c_{11}(X)} (\sqrt{1 + \frac{c_{12}^2(X)}{c_{11}^2(X)}} - 1),
\]

(5.25)
which is the optimal parameter for $Q(x; \theta_i)$. Indeed, substituting the solution (5.25) in the second derivative of $Q(x; \theta_i)$ with respect to $p$ we have:

$$\frac{d^2}{dp^2} Q(x; \theta_i) = \frac{-4c_{i1}(X)\hat{p} \left( \frac{c_{i2}^2(X)}{c_{i1}^2(X)} - 1 \right)}{(1 - \hat{p}^2)^3} \geq 0,$$

where $\hat{\theta}_i = (p_0, p_1, \ldots, p_{l-1}, \hat{p}, \hat{p}, \ldots)$.

**Lemma 5.2.7** In the Kautz series expansions, the optimum parameter for the enforced convergence criterion $Q(x; \theta_i)$, where the first $l$ $p_k$’s are taken fixed, is given by

$$\hat{p} = \frac{c_{i2}(X)}{c_{i1}(X)} \left( \sqrt{1 - \frac{c_{i1}^2(X)}{c_{i2}^2(X)}} - 1 \right),$$

where $x \in \mathcal{D}_d^*$. The minimum enforced convergence criterion is

$$Q(x; \hat{\theta}_i) = -\frac{1}{2}c_{i1}(X) \sqrt{1 - \frac{c_{i1}^2(X)}{c_{i2}^2(X)}} \left[ 1 + \sqrt{1 - \frac{c_{i1}^2(X)}{c_{i2}^2(X)}} \right] \left( c_{i3}(X) \right).$$

where $\hat{\theta}_i = (p_0, p_1, \ldots, p_{l-1}, \hat{p}, \hat{p}, \ldots)$.

**Theorem 5.2.8** Let $x$ be a non zero function in the admissible space $\mathcal{D}_d$, and $\theta_i(p) = (p_0, p_1, \ldots, p_{l-1}, p, p, \ldots)$, where $p, p_k \in (-1, 1)$, $k = 0, 1, \ldots, l - 1$, $l \in \mathbb{N}$, and $p_k$’s are fixed. Further, let $\theta_{i+1}(p) = (p_0, p_1, \ldots, p_{l-1}, p, p, \ldots)$, where $p_i = \hat{p}$ is the optimal solution of $\min_{p \in [-1, 1]} Q(x; \theta_i(p))$. Then

$$\min_{p \in [-1, 1]} Q(x; \theta_{i+1}) \leq \min_{p \in [-1, 1]} Q(x; \theta_i).$$

**Proof:** Let $\hat{p}$ be the optimal solution for $\min_{p \in (-1, 1)} Q(x; \theta_i)$, then

$$Q(x; \theta_i(\hat{p})) = \min_{p \in [-1, 1]} Q(x; \theta_i),$$

if $\theta_i(\hat{p}) = (p_0, p_1, \ldots, p_{l-1}, \hat{p}, \hat{p}, \ldots)$, so for the minimum

$$\min_{p \in [-1, 1]} Q(x; \theta_{i+1}) \leq \min_{p \in [-1, 1]} Q(x; \theta_i).$$

For the recursive solution of the real Kautz system we have the following algorithm:

---
Algorithm 5.2.9 For $x \in D_0^*$:

1. $N \in \mathbb{N}$;

2. $p_0$: optimal Laguerre solution;

3. do $l = 1; N - 1$,
   \[ \beta := \frac{c_{2l}(X)}{c_{2l}(X)}; \]
   \[ p_l := \beta \left( \sqrt{1 - \frac{1}{\beta^2}} - 1 \right); \]
   od;

\[ \square \]

Example 5.2.10 As an example, we take the function $x$ defined by a rational $z$-transform:

\[ X(z) = \frac{\prod_{i=1}^{3} (z - q(i))}{\prod_{i=1}^{4} (z - p(i))}, \]

with

\[ q = (-1, 1, 1), \]
\[ p = (0.3, -0.3, 0.4, -0.4). \]

The function $x$ is real-valued. In Fig 5.6 the amplitude spectrum is given. We observe a band-pass filter. The optimal Laguerre solution is given by $\hat{\rho} = -0.1833$. Applying the algorithm given above to this function for $N = 6$ we obtain for the optimal first 6 parameters:

\[ \hat{\rho}^{1-6} = (-0.1833, -0.1567, -0.2728, 0.4526, 0.1091, -0.262). \]

![Figure 5.6: Amplitude spectrum of the function $x$.](image-url)
5.2 Optimal parameters in Kautz series expansions

In Fig 5.7, the relative value of $Q(x; \theta) (= Q(x; \theta)/\|x\|^2)$ is shown as a function of the number of iterations. The relative enforced criterion is decreasing as function of the number of iterations. We see that $Q(x; \theta)$ becomes quite constant after 6 iterations. In Fig 5.8, the relative energy loss of $x$ is shown as a function of the number of sections in the Kautz series (solid line). We also calculated the loss of energy for a truncated Laguerre series where the pole was optimized ($\bar{p} = -0.1833$) according to the enforced convergence criterion for the discrete Laguerre case (dashed line). We see that the Kautz series is better than the Laguerre series in this case. We note that this example is representative for functions taken in span$\{\lambda_0^t, \lambda_1^t, \ldots, \lambda_m^t\}$.

Figure 5.7: Enforced convergence criterion as function of the iteration.

Figure 5.8: Relative energy loss in the Kautz series with one optimal pole (dashed line) and recursively with different poles (solid line).
5.2.6 Highly oscillating systems

The complex Kautz series expansions can be treated in the same way as the real Kautz series expansions. We restricted our attention to the real case in the previous subsection for reasons of simplicity. For a real function with a clearly resonant character we combine the two previous algorithms. By making use of the equations (5.16), (5.21), (5.22) and (5.23), in the complex pole case, we obtain the complex version of equation (5.24):

\[ Q(x; \theta) = \frac{1}{2} \bar{c}_{11}(X) \frac{p + p^*}{1 - pp^*} + \bar{c}_{22}(X) \frac{1}{1 - pp^*} + \bar{c}_{13}(X), \]

where \( \bar{c} \)'s are identical to the \( c \)'s defined earlier but where conjugations have been added to the appropriate places. In the following, we give an algorithm for the recursive solution of highly oscillating systems.

Algorithm 5.2.11 For \( x \in \mathcal{D}_d^* \):

1. \( N \in \mathbb{N} \);
2. \( y = \text{hilbert}(x) \);
3. \( p_0 \): optimal complex Laguerre parameter for \( y \);
4. do \( l = 1 : N - 1 \),
   \[ \beta := \frac{\bar{c}_{11}(X)}{\bar{c}_{11}(X)}; \]
   \[ p_l := \beta^* \left( \sqrt{1 - \frac{1}{pp^*}} - 1 \right); \]
   od;
5. \( \hat{\theta} = (p_0, p_0^*, p_1, p_1^*, \ldots) \);

\[
\blacksquare
\]

5.3 Discussion

We have considered functions having finite energy and finite center of the energy expanded in a modulated Laguerre and Kautz series expansions. In order to obtain a compact series expansion, we introduced an enforced convergence criterion. The chosen enforced convergence criterion equals the first-order moment of the energy distribution in the transform domain. This specific criterion, and its minimization, is especially suitable for the considered series expansion since it allows us to make use of the differential equation associated with the modulated Laguerre functions, and the recursive differential equation associated with the Kautz functions.
5.3 Discussion

Laguerre expansions

We used the free parameters in the modulated Laguerre series to minimize this criterion. As a consequence, simple expressions for the optimum compaction parameters evolve. The optimum free parameters have been established in this chapter, and an upper bound for the quadratic truncation error was given. By the Moebius transformation we have shown the equivalence between the discrete-time and the continuous-time cases to determine the optimal parameters.

Kautz expansions

We proposed a simple procedure to establish the optimal parameter in a Kautz series expansion characterized by repeated complex conjugated pole-pair on the basis of an energy compaction criterion. This criterion is suboptimal in a quadratic sense, but yields simple explicit expressions for the optimal pole in terms of the Hilbert transform of the given function. In contrast to an optimal pole according to a quadratic error criterion, the enforced convergence criterion has the convenient property that it defines a pole independent of the number of the terms in the expansion. By an example, it was shown that such procedure can give results close to the optimal one in a quadratic sense.

A second approach to minimize the enforced convergence criterion for the general Kautz systems was presented by making use of the properties of the recursive form of the differential equation associated with the Kautz functions. By an algorithm, we have shown how to determine the optimal parameters in the recursive scheme. Combinations of the first and second method can be taken, which leads to Algorithm 5.2.11. By an example we made a comparison between the Laguerre solution and the recursive Kautz model, to show how the recursive Kautz solution performs.

Finally, we note that the optimal pole defined by the enforced convergence criterion can always be used as the initial estimate in a more refined optimization procedure, for example, in the quadratic error sense.
Part II

Filter banks in a design perspective
Chapter 6

Introduction to Part II

The key words in this part are filter bank, localizing unitary transformation and frame. They all play important roles in signal processing. A filter bank, as depicted in Fig. 6.1, is a part of source-coding systems for audio and video as used in electronic industry. They are present in different fields in applied science, such as digital audio, speech and image compression, and statistical signal processing. In contrast, localizing unitary transformations are more involved in the theoretical aspects of signal processing. Frames can be seen as the bridge between them. From the one side, the redundant information of the localized transform can be reduced to 'over'-complete systems (frames) and, from the other side, the system associated with the analysis/synthesis filter bank should constitute a frame, in order to guarantee a perfect reconstruction filter bank.

Filter banks also represent analysis and synthesis of the signals. They fit in special classes of localizing unitary transformations. The most popular filter banks of this type are the ones obtained by the short-time Fourier\(^1\) transformation and the wavelet transformation. The localizing unitary transformations analyze the instantaneous nature of the signals. The operation $\text{signal} \mapsto \text{localized transform}$, is considered as a filtering operation. Thus an input - output system is defined where the input is the signal and the output is its localized transform. Typically, the localized transform yields a redundant representation since the localizing unitary transformation extends the dimension of the signal variable. This can be used to reduce the localized transform, that is, localized spectrum, to a certain grid, resulting in a sampled localized spectrum. The operation $\text{signal} \mapsto \text{localized transform} \mapsto \text{sampled localized transform}$ can be regarded as the analysis part of the filter bank. If it is possible to reconstruct the signal from its sampled local spectrum then the reconstruction scheme is the right hand side or the synthesis part of the filter bank.

\(^1\)Joseph Fourier lived from 1768 to 1830: attended the Ecole Royale Militaire of Auxerre. In 1794 Fourier was nominated to study at the Ecole Normale in Paris. The school opened in January 1795 and Fourier was certainly the most able of the pupils whose abilities ranged widely. By 1 September 1795 Fourier was a teacher at the Ecole Polytechnique.

By 1807 Joseph Fourier had completed his important memoir On the Propagation of Heat in Solid Bodies.
An analysis filter bank or a sampled local spectrum is a filtering scheme, where
the output is the result from taking the inner product of the signal with the collection
of analyzing signals identified by the local transformation and the grid. This collection
 guarantees a reconstruction of the signal if this collection constitutes a frame in
the signal space.

6.1 Uniformly downsampling filter bank

Given an analysis filter bank scheme, as in Fig. 6.2, an input-output system is defined,
where \( x_{in} \) is an input signal and \( y_{out} \) is the output of the analysis filter bank.
Mathematically, the transfer from \( x_{in} \) to \( y_{out} \) is given by

\[
y_{out}(l) = \sum_{k} x_{in}(k) A_m(lL - k) = \langle x_{in}, \sigma^{lL} \tilde{f}_{m} \rangle, \quad l \in \mathbb{Z}, m = 0, \ldots, M - 1,
\]

where \( \sigma \) is the right-shift operator and * denotes time reversal. From this representation, we can derive necessary and sufficient conditions for perfect reconstructing
filter banks. A reconstruction scheme is given by the synthesis filter bank drawn
in Fig. 6.3. The well-known filter banks are those related to the Sliding-Window
Discrete-Time Fourier Transformation (SWDTFT). The relationship between the filter
bank and those transformations is obvious: applying a localizing unitary transform-
ation \( \mathcal{W} \) to the signal \( x_{in} \) results in the localized transform \( \mathcal{W} x_{in} \), which is a
function of two variables, for instance, \((l, \omega)\). Restricting the localized transform

Figure 6.1: Filter bank.
6.1 Uniformly downsampled filter bank

$\mathcal{W}x_{in}$ to a grid $\Gamma \subseteq \mathbb{Z} \times \mathbb{R}$ defined by $\Gamma = \{(lL, \omega_m)| l \in \mathbb{Z}, m = 0, \ldots, M - 1\}$, the redundancy of the localized transform is reduced where for the grid $\Gamma$ we have:

$$(\mathcal{W}x_{in}) (lL, \omega_m) = \langle x_{in}, e^{jlL\omega_m} \rangle, \quad l \in \mathbb{Z}, m = 0, \ldots, M - 1,$$

where $\Psi_m$ is identified by the kernel of the localizing unitary transformation $\mathcal{W}$. So it seems that there is a strong relationship between the concept of filter bank and localizing unitary transformation. To come to a filter bank we remark that there is freedom in choosing the localizing unitary transformation, that is, not necessarily the short-time Fourier transformation or the wavelet transformation but, for instance, transformations depending on the application where the filter bank is used.

The short-time Fourier transformation is defined as the cross-ambiguity function of the signal and a window function, and is constructed in the following way. The signal is multiplied by translated versions of a window, which is usually more or less concentrated around a certain time moment, and the discrete-time Fourier transform of the product is determined. Thus, a function of time and frequency is created, which may be considered the local frequency spectrum of the signal. The concept of SWDTFT can be unified by replacing the kernel with an arbitrary kernel-function of a unitary transformation, leading to the so-called Windowed Unitary Transformation (WUT) [18, 76, 78].

It is well known that a spectrum of the SWDTFT is completely determined by its values on the points of a certain time-frequency lattice, which is exactly the lattice suggested by Gabor in the SWDTFT case [5]. In the same way, the WUT is a redundant description of the signal. Thus only partial knowledge of the WUT is sufficient

---

Dennis Gabor lived from 1900 to 1979: entered the Technische Hochschule Berlin and acquired a Diploma in 1924 and his Doctorate of Engineering in 1927 in electrical engineering. On January 1, 1949 he joined the Imperial College of Science & Technology in London, first as a Reader in Electronics, and later as Professor of Applied Electron Physics, until his retirement in 1967. He tackled many problems from communication theory, the elucidation of Lamuirs Paradox, the inexplicably intense apparent electron interaction in low pressure mercury arcs.
to reconstruct the signal. Of course the question arises: What are the conditions (necessary and sufficient) on a lattice such that knowledge of the WUT on this lattice is sufficient for complete reconstruction.

Given a unitary transformation and a causal window function the Causal Sliding Windowed Unitary Transformation (CSWUT) is introduced in Chapter 7. The theory of frames [9] yields necessary conditions for perfect reconstruction given a subsampling of the local spectrum of a signal.

In practice, analysis and synthesis filter banks come into play to implement a partial description and algorithms to reconstruct the signal from this partial knowledge. Given an analysis filter bank, simple causal reconstruction schemes (in terms of necessary filtering operations) are desired. In Section 7.6 of Chapter 7, we treat two examples namely the Windowed Laguerre and Kautz Transformations.

### 6.2 Non-uniformly downsampled filter bank

An efficient subsampling of a localized spectrum is not necessarily uniform. In order to be able to tailor the filter bank to the application at hand, it is desirable to have a theory as general as possible. Therefore, rather arbitrary analysis filter banks (causal and stable) have to be considered, see Fig. 6.4, followed by arbitrary downsamplers. In this way we have the desired flexibility in choice of the filter bank and the possibility to deal with issues such as signal quantization noise and coefficient quantization.

While on the one hand one requires maximal flexibility in design, it is often desired to have insight in the representation of the signal as is given by the output of the analysis bank. This is for instance the case if the filter bank implements a subsampled sliding-window discrete-time Fourier transform or the analysis agrees with a wavelet representation or multi-resolution. A part of this chapter has been published in [77].
6.3 Goals

The goals of this part are summarized in the following items:

- Starting from the concept of analysis filter bank and knowing their subsampling properties, determine a class of localizing transformations that can be used to build up perfect reconstruction filter banks.

- Show the relations between filter banks, localizing transformations and frames.

- Given a uniformly subsampled localized transform, give simple reconstruction schemes for the synthesis filter bank.

- Formulate the problem of non-uniform subsampling, or non-uniformly downsampled filter banks, give necessary and sufficient condition to perfect reconstruction, and develop reconstruction schemes.

- Give illustrations using simple (low-order) filters, in particular, the Kautz and Laguerre filters.

6.4 Outline

In Chapter 7, we provide the uniformly downsampled case. We present the basic building blocks in a filter bank, and give their properties. We deal with different concepts that play a role in a multirate system, like causality and stability. Filter banks...
recognize different building structures, some of which are efficient in implementation. We present these structures and deal with the concept of perfect reconstruction. We clarify the equivalence between filter banks and frames. The local aspect of filter banks can be comprehended in localizing unitary transformations. We deal with a special type of localizing unitary transformations and its relation to filter banks and frames. Further, we give reconstruction schemes by making use of linear constant matrix coefficient difference equations. As illustrations we consider the Laguerre and Kautz filters to clarify the theory presented in the chapter.

In Chapter 8, we consider the non-uniformly downsampled case. We define the reconstruction problem, we reformulate it to subsampled localized transforms and frames and we give conditions for perfect reconstruction. We create a uniform down-sampled analysis filter equivalent to the non-uniform one, and reconstruction schemes by making use of the schemes described in Chapter 7.

“Yesterday was my 21st birthday, at that age Newton and Pascal had already acquired many claims to immortality.”

A citation from a letter of Fourier
Chapter 7

Uniformly downsampled filter banks

The purpose of this chapter is to provide the local analysis of signals by filter banks, induced by localizing transformations. We give the preliminaries needed to deal with filter banks, and consider local analysis by the following three items:

1. Analysis by a filter bank as a set of analyzing filters in parallel followed by a compression factor, which can be seen as a system with single input and multiple output.

2. Analysis by a frame, that is, extracting the characteristics of a signal by taking it in the inner product with the elements of an ‘over’-complete system.

3. Analysis by a localized unitary transform, that is, applying a unitary transformation of a localized signal by a window, resulting in a localized spectrum.

As we will see in this chapter, the three items described here show equivalences in the treatment of signals. Moreover, it is sometimes useful to consider the properties of one of the three items to derive properties for another one. For example, determining the frame bounds required by item 2 can be done by determining the extremes of the eigenvalues of a matrix obtained from the analysis polyphase matrix related to item 1, see Section 7.4.

The setup of this chapter is as follows. In the first section of this chapter, we present the basic building blocks in a filter bank, and give their properties. Further, we deal with different concepts that play a role in a multirate system, like causality and stability. Filter banks recognize different building structures, some of which are efficient in implementation. In Section 7.2, we present these structures and deal with the concept of perfect reconstruction. As mentioned before, filter banks show similarity with frames. In Section 7.3, we discuss the equivalence between filter banks and frames. The local aspect of filter banks can be comprehended in localizing unitary transformations. In Section 7.4, we deal with a special type of localizing unitary transformations and its relation to filter banks and frames. In Section 7.5, we give reconstruction schemes by making use of linear constant matrix coefficient difference equations. In Section 7.6, illustrations by Laguerre and Kautz filters clarify the
theory presented in Section 7.4 and Section 7.5. The Laguerre and Kautz filter banks yield a simple filtering scheme of perfect reconstructing filter bank. We conclude the Chapter by conclusions in Section 7.7.

7.1 Basic building blocks in filter banks

In this thesis, a filter $H$ is a linear time invariant input-output system, see Fig. 7.1, described by the relationship

$$y = h * x,$$

where $x$ is the input, $y$ the output and $h$ the impulse response of the filter $H$.

![Figure 7.1: Input-output scheme for the filter $H$.](image)

Recalling that the Discrete-Time Fourier Transformation (DTFT) $\mathcal{F}_D$ of the convolution of two signals is the product of their transforms, we write

$$Y(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega}),$$

where the discrete-time Fourier transformation, $\mathcal{F}_D : x \rightarrow X$, is defined by:

$$X(e^{j\omega}) = \sum_{t \in \mathbb{Z}} x(t)e^{-j\omega t}, \quad \omega \in (-\pi, \pi], \quad (7.1)$$

for signals $x$ with $\sum_{t \in \mathbb{Z}} |x(t)| < \infty$. The inversion from $X$ to $x$ is obtained by:

$$x(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega})e^{j\omega t}d\omega, \quad t \in \mathbb{Z}.$$

The function $H(e^{j\omega})$ is called the frequency response of $H$ and will be identified with the filter $H$. The $z$-transformation $\mathcal{Z}$, which is a variant of the discrete-time Fourier transformation, is defined by:

$$X(z) = (\mathcal{Z}x)(z) = \sum_{t \in \mathbb{Z}} x(t)z^{-t}. \quad (7.2)$$

The inversion from the $z$-transform of $x$ is obtained by:

$$x(t) = \frac{1}{2\pi j} \oint_{C} X(z)z^{-t-1}dz,$$
with \( C \) is a contour encircling \( z = 0 \), in the region where \( X \) is analytic. We use for the discrete-time Fourier transform and the \( z \)-transform the same notation \( X \), since there is no confusion in considering the two at the same time. So by the \( z \)-transformation we have also

\[
Y(z) = H(z)X(z)
\]

Note that the \( z \)-transformation is the discrete-time Fourier transformation if we evaluate at \( z = e^{j\omega} \). The \( z \)-transformation maps the signal (sequence) to a function defined on the complex plane, while the discrete-time Fourier transformation maps the signal to a function defined on the unit circle, as a function of \( e^{j\omega} \), or to a function defined on \([−\pi, \pi]\) as a function of \( \omega \). The relation between the discrete-time Fourier transformation and the \( z \)-transformation is used in a similar way as the relation between the Fourier transformation and the Laplace transformation in the continuous-time case. Of course, the advantage of the use of the \( z \)-transformation, c.q. Laplace transformation in continuous-time, is related to regions of convergence and stability. For example, for the sequence \( h(t) = a^t, t \geq 0 \), and \( h(t) = 0 \) for \( t \leq 0 \), it is obvious that the \( z \)-transform is \( H(z) = 1/(1 - az^{-1}) \), and that the summation in (7.2) converges if \( |z| > |a| \). The region identified by \( |z| > |a| \) is called the region of convergence of \( H(z) \). The region of convergence is defined to be the part in the complex plane where the series associated to \( H(z) \), see (7.2), is convergent. For a rational transfer function \( H(z) \), the region of convergence is bounded by the smallest circle which contains all the poles. A filter is called (BIBO) stable if any bounded input signal, leads to a bounded output signal. It can be shown that the filter, with rational transfer function, is stable if and only if its region of convergence contains the closed unit circle \( U \).

A filter \( H \) is said to be causal if the output at time \( t = t_1 \) of the system \( H \) does not depend on future values \( t > t_1 \) of the input. Mathematically, a filter \( H \), with impulse response \( h \), is causal if and only if, for any inputs \( x_1, x_2 \) and \( t_1 \in \mathbb{Z} \):

\[
x_1(t) = x_2(t), \quad \text{for} \ t \leq t_1,
\]

implies

\[
(h * x_1)(t) = (h * x_2)(t), \quad \text{for} \ t \leq t_1.
\]

Obviously, a filter \( H \) is causal if and only if the support of its impulse response is a part of \( \mathbb{Z}_0^+ \). We give the definition of the two types of filters: FIR filters which are filters with finite duration impulse responses, and IIR filters which are filters with infinite duration impulse responses. For a mathematical definition of realizable linear time-invariant filters we consider the polynomials. A polynomial \( p \) over \( \mathbb{C} \) in the variable \( z \) is an expression of the form

\[
p(z) = \sum_{0}^{d} a_i z^{i}
\]
where $0 \leq d$ are integers and $a_i$ are elements of $\mathbb{C}$. The set of all polynomials is denoted by $\mathbb{C}[z]$. A rational function $r$ over $\mathbb{C}$ in the variable $z$ is an expression of the form

$$r(z) = q(z)^{-1}p(z)$$

where $p, q \in \mathbb{C}[z]$. The set of all rational functions is denoted by $\mathbb{C}(z)$. The set $\mathbb{C}[z]$ is a ring without zero divisors with quotient field $\mathbb{C}(z)$. A filter $H(z)$ is an FIR filter if $z^N H(z) \in \mathbb{C}[z]$, for some integer $N$. A filter $H(z)$ is an IIR filter if it is an element in $\mathbb{C}(z)$ but not FIR.

A transfer function $H(e^{j\omega})$ can be factorized as follows:

$$H(e^{j\omega}) = |H(e^{j\omega})|e^{j\arg H(e^{j\omega})},$$

where $|H(e^{j\omega})|$ represents the amplitude spectrum of the filter $H$, and $\arg(H(e^{j\omega}))$ the phase. This representation may help to understand the bandpass character of the filter. A bandpass filter is a filter that passes one frequency band and attenuates frequencies above and below that band. The bandwidth of $H(z)$ is the frequency width of the passband of a filter. For a low-pass filter, the bandwidth is equal to the cutoff frequency. For a bandpass filter, the bandwidth is typically defined as the frequency difference between the upper and lower -3 dB points. An ideal low pass filter has the frequency response:

$$H(e^{j\omega}) = \begin{cases} e^{-j\omega t_0} & \text{if } |\omega| < \omega_0, \\ 0 & \text{otherwise.} \end{cases}$$

where $t_0 \in \mathbb{Z}$ and $0 < \omega_0 < \pi$. The filter $H$ is ideal because its amplitude is the characteristic function of an interval. Its impulse response is given by:

$$h(t) = \frac{\sin \omega_0 (t - t_0)}{\pi (t - t_0)}.$$  

We note that the filter $H$ is not rational. In engineering practice, the following way is used to factorize filters:

$$H(e^{j\omega}) = e^{-j\phi(\omega)}R(\omega).$$

In the following, we consider the down- and upsampling. A downsampler is also called decimator or compressor. In Fig. 7.2 we show the block diagram of the $L$-fold downsampler.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{L-fold-downsampler.png}
\caption{The $L$-fold downsampler.}
\end{figure}
7.1 Basic building blocks in filter banks

Definition 7.1.1 The $L$-fold downsampler ($\downarrow L$) is defined by

1. 
   \[(\downarrow Lx)(t) = x(tL), \ t \in \mathbb{Z},\]
   or, in terms of the $z$-domain,
2. 
   \[(\downarrow Lx)(z^L) = \frac{1}{L} \sum_{i=0}^{L-1} X(\Omega_L^i z).\]

where $\Omega_L = e^{-j2\pi/L}$, for a discrete-time signal $x$. □

Only the input sample numbers equal to a multiple of $L$ are retained. We say, there is aliasing if the supports of $X(\Omega_L^i e^{j2\pi\omega})$, $i = 0, \ldots, L - 1$, are overlapping for $\omega \in [-\pi, \pi]$. In general downsampling introduces aliasing, except for special cases of bandlimited signals. So in general it is not possible to recover the signal $x$ from $\downarrow Lx$. In practice, the downsampler is preceded by a frequency band-pass (low-pass) filter to prevent or reduce the aliasing caused by the downsampler. The filter ensures that the signal being downsampled is approximately bandlimited. The exact band-edges of the filter depend on how much aliasing is permitted.

In Fig. 7.3 we show the block diagram of the $L$-fold upsampler, also called expander or interpolator.

Definition 7.1.2 The $L$-fold upsampler ($\uparrow L$) is defined by

1. 
   \[(\uparrow Lx)(t) = \begin{cases} 
   x(\frac{t}{L}) & \text{if } t \mod L = 0 \\
   0 & \text{otherwise} \end{cases},\]
   or, in terms of the $z$-domain,
2. 
   \[(\uparrow Lx)(z_L) = X(z^L),\]

for a discrete-time signal $x$. □

\[\xymatrix{ x \ar[r]^L & y }\]

Figure 7.3: The $L$-fold upsampler.
Uniformly downsampled filter banks

The output of the $L$-fold upsampler is obtained by inserting $L - 1$ zeros between two adjacent samples of the input signal. It is clear that in this case the input signal $x$ can be recovered from $\uparrow Lx$, since there is no loss of information. The upsampler is followed by a low-pass filter with cut-off frequency $\pi / L$, in order to suppress the images created by the upsampler. We note that the down- and upsampler are non-causal time-varying linear systems, in a strict mathematical sense. Further, we have $\downarrow L \uparrow L = I$.

A uniformly downsampled filter bank is a collection of filters operating in parallel, either in an analysis mode with a single input and multiple outputs or in a synthesis mode with multiple inputs and a single output. An analysis filter bank, as shown in Fig. 7.4, is used to decompose a signal into $M$ components, the so-called subband signals $y_{out0}, y_{out1}, \ldots, y_{outM-1}$. The building blocks are the filters $A_0, A_1, \ldots, A_{M-1}$, the so-called analysis filters, and the downsampling factor $L$.

The filter bank is called uniformly downsampled because we use the same downsampling factor in every channel. In a more general situation, downsampling may differ from channel to channel, depending on the characteristics of the analysis filters. In this case we call the filter bank non-uniformly downsampled.

**Figure 7.4:** Analysis filter bank.

**Figure 7.5:** Synthesis filter bank.
The synthesis filter bank, as shown in Fig. 7.5, enables the merging of the signals $y_{\text{out}}, y_{\text{out} 1}, \ldots, y_{\text{out} M-1}$ in order to reconstruct the input signal $x_{\text{in}}$. It consists of an upsampling factor $L$, and synthesis filters $S_0, S_1, \ldots, S_{M-1}$. Adding the output signals of the synthesis filters, the output signal $x_{\text{out}}$ of the synthesis filter bank is obtained. Filter banks have to be designed such that $x_{\text{out}} \approx x_{\text{in}}$.

Next we introduce the noble identities which are useful in the theory and implementation of filter banks, where we may reverse the order in a filter bank, as shown in Fig. 7.6, of the filters and the down- or upsamplers.

- Downsampling by $L$ followed by filtering with $H(z)$ is equivalent to filtering with the upsampled filter $H(z^L)$ before the downsampling.
- Upsampling by $L$ followed by filtering with upsampled $H(z^L)$ is equivalent to filtering with the filter $H(z)$ before the upsampling.

### 7.2 Structures and perfect reconstruction in filter banks

Signal processing by filter banks has become more attractive since the invention of the polyphase representation [7, 89]. This representation enables the user to simplify the computations of the theoretical results, and leads to efficient implementations of filter banks. To explain the basic idea, let $H(z)$ be a filter with impulse response $h$. The type-1 polyphase representation of the filter $H(z)$ reads

$$H(z) = \sum_{l=0}^{L-1} z^{-l} E_l(z^L), \quad (7.3)$$

The schematic representation of the relation between $h$ and the impulse response $e_l$ of the $l$th polyphase component.
with
\[ E_l(z) = \sum_{k \in \mathbb{Z}} h(l + kL)z^{-k}. \]

where \( E_l(z) \) is the transfer function of the \( l \)th polyphase filter, also called polyphase component. In Fig. 7.7, we summarize the \( l \)th polyphase component in a scheme. The \( M \times L \) analysis polyphase matrix \( \mathbf{E}(z) \) of the analysis filter bank, as shown in Fig. 7.4, is defined as
\[ \mathbf{E}(z)(m, l) = E_{m,l}(z), \ m = 0, \ldots, M - 1, \ l = 0, \ldots, L - 1, \quad (7.4) \]

where \( E_{m,l}(z) \) is the \( l \)-th type-1 polyphase component of the \( m \)-th filter \( A_m \):
\[ E_{m,l}(z) = \sum_{k \in \mathbb{Z}} a_m(l + kL)z^{-k}, \]

here \( a_m \) is the impulse response of \( A_m \).

The type-2 polyphase representation of \( H(z) \) is a variation of (7.3):
\[ H(z) = \sum_{l=0}^{L-1} z^{-L-1-l} R_l(z^{-L}), \quad (7.5) \]

where \( R_l(z) \) are permutations of \( E_l(z) \):
\[ R_l(z) = E_{L-1-l}(z). \]

the functions \( R_l(z) \) are called the type-2 polyphase components. The \( L \times M \) synthesis polyphase matrix \( \mathbf{R}(z) \) of the synthesis filter bank, as shown in Fig. 7.5, is defined as
\[ \mathbf{R}(z)(l, m) = R_{m,l}(z), \ l = 0, \ldots, L - 1, \ m = 0, \ldots, M - 1, \quad (7.6) \]
7.2 Structures and perfect reconstruction in filter banks

Figure 7.9: Efficient structure of a synthesis filter.

where $R_{m,l}(z)$ is the $l$-th type-2 polyphase component of the $m$-th synthesis filter $S_m$:

$$R_{m,l}(z) = \sum_{k \in \mathbb{Z}} s_m(L - 1 - l + kL)z^{-k},$$

here $s_m$ is the impulse response of $S_m$.

Type-1/type-2 polyphase representation is useful for efficient implementation of analysis/synthesis filter banks. For an analysis filter $H(z)$ with polyphase representation (7.3), we can redraw the scheme in the left-hand side of Fig. 7.8. By invoking the noble identity for the downsampler, by interchanging of filtering and downsampling, this can be redrawn as in the right-hand side of Fig. 7.8. This implementation is more efficient than a direct implementation of $H(z)$. Indeed, by counting the additions and multiplications we see the improvement in efficiency: Suppose $L = 2$, and $H(z)$ is an FIR filter, requiring $N$ operations per unit time (input sample spacing), operating at 10kHz, for a total of $10^4 N$ operations per second. Then $E_0(z)$ and $E_1(z)$ require $\frac{N}{2}$ operations per unit and operating at 5kHz. So the total number of operations per second is $2 \times \frac{N}{2} \times 5 \times 10^3 = 5 \times 10^3 N$, that is, the polyphase implementation requires half the number of operations per second of the filter $H(z)$. We note that the implementation described here is done only for one channel in an analysis filter bank.

For synthesis filter banks we can do the same. A more efficient structure, see Fig. 7.9, can again be obtained by using the type-2 polyphase representation. In the left-hand side we show the type-2 polyphase representation, and in the right-hand side the polyphase components at lower rate (moved before the upsampler). By using the efficient structures of analysis and synthesis filter banks, we can redraw the filter bank in the equivalent form shown in Fig. 7.10. This simplified structure can even
be used in practical implementations, and has the advantage that the filters \( \mathbf{E}(z) \) and \( \mathbf{R}(z) \) are operating at the lower rate [89].

In many applications the analysis filter bank, see Fig. 7.4, and the synthesis filter bank, see Fig. 7.5 are connected back to back, that is, the tuple of vectors \( \mathbf{y}_{\text{in}} \) is taken to be equal to the tuple of vectors \( \mathbf{y}_{\text{out}} \). In that case, the expression for the input-output transfer of the filter bank is given by:

\[
X_{\text{out}}(z) = \frac{1}{L} \sum_{m=0}^{M-1} \sum_{l=0}^{L-1} S_m(z) A_m(\Omega_L^l z) X_{\text{in}}(\Omega_L^l z)
\]

\[
= \frac{1}{L} \left( \sum_{m=0}^{M-1} S_m(z) A_m(z) \right) X_{\text{in}}(z)
\]

\[
+ \frac{1}{L} \sum_{m=0}^{M-1} \sum_{l=1}^{L-1} S_m(z) A_m(\Omega_L^l z) X_{\text{in}}(\Omega_L^l z),
\]  

(7.7)

The term \( \mathbf{A} \mathbf{T}(z) := \frac{1}{L} \sum_{m=0}^{M-1} \sum_{l=1}^{L-1} S_m(z) A_m(\Omega_L^l z) X_{\text{in}}(\Omega_L^l z) \) is called the alias term. In matrix notation, the input-output relation of the filter bank is given by:

\[
X_{\text{out}}(z) = \frac{1}{L} \mathbf{X}_{\text{in}}(z) \mathbf{A}(z) \mathbf{S}(z), \ z \in \mathbb{U}
\]  

(7.9)
7.2 Structures and perfect reconstruction in filter banks

where

\[
X_{in}(z) = \left( X_{in}(z), X_{in}(\Omega_L z), \ldots, X_{in}(\Omega_L^{L-1} z) \right),
\]

\[
S(z) = (S_0(z), S_1(z), \ldots, S_{M-1}(z))^\top,
\]

\[
A(z) = \begin{pmatrix}
A_0(z) & A_1(z) & \ldots & A_{M-1}(z) \\
A_0(\Omega_L z) & A_1(\Omega_L z) & \ldots & A_{M-1}(\Omega_L z) \\
\vdots & \vdots & \ddots & \vdots \\
A_0(\Omega_L^{L-1} z) & A_1(\Omega_L^{L-1} z) & \ldots & A_{M-1}(\Omega_L^{L-1} z)
\end{pmatrix}.
\]

The matrix \( A(z) \) is called the alias component matrix. The alias component matrix \( A(z) \) and the polyphase component matrix of an analysis filter bank are related as follows [89]:

\[
A(z) = W^* D(z) E^\top (z^L),
\]

where \( W \) is the DFT \( L \times L \) matrix:

\[
W(k, l) = \Omega_L^k, \quad l, k = 0, \ldots, L - 1, \quad \Omega_L = e^{-j\frac{2\pi}{L}},
\]

and \( D(z) \) is the delay diagonal \( L \times L \) matrix:

\[
D(z) = \text{diag}(1, z^{-1}, \ldots, z^{-(L-1)}).
\]

The synthesis vector \( S(z) \) is related to the polyphase component matrix \( R(z) \) of a synthesis filter bank according to:

\[
S(z) = R^\top (z^L) V(z),
\]

where

\[
V(z) = \left( z^{-(L-1)}, z^{-(L-2)}, \ldots, 1 \right)^\top.
\]

The input in the representation (7.9) is the vector function obtained from the modulated versions of \( X_{in}(z) \), \( X_{in}(\Omega_L z) \), \( l = 0, \ldots, L-1 \). We say that the filter bank is alias-free if the alias term \( A(z) T(z) \) is zero, \( z \in U \), for all admissible input signals \( X_{in}(z) \), which means that the filter bank defines a linear time invariant system. For special cases of the synthesis filters \( S_m(z) \), \( m = 0, \ldots, M - 1 \), that is, satisfying the conditions:

\[
\sum_{m=0}^{M-1} S_m(z) A_m(\Omega_L^l z) = 0, \quad l = 1, \ldots, L - 1,
\]

(7.14)
the aliasing is completely canceled in the input-output transfer function. For the necessary and sufficient condition for alias cancellation we refer to [89, 90]. So aliasing is permitted in the analysis filter bank and completely canceled through the synthesis bank. This means, it is not necessary to avoid aliasing in the analysis bank if it is possible to cancel it by the synthesis filter bank. For example, if the analysis filters $A_m(z), m = 0, \ldots, M - 1,$ have large transition bandwidths and low stopband attenuations, which means large aliasing errors, then the errors can be canceled by choosing synthesis filters satisfying (7.14).

For alias free filter banks we have the input-output relation:

$$X_{\text{out}}(z) = T(z)X_{\text{in}}(z), \ z \in \mathbb{U}.$$  

The function $T(e^{j\omega})$ is called the distortion transfer function [89], also called overall transfer function. If the function $|T(e^{j\omega})|$ is not a non-zero constant, we say that $X_{\text{out}}(e^{j\omega})$ suffers from amplitude distortion ($|T(e^{j\omega})|$). Similarly, if $\arg T(e^{j\omega})$ is not linear in $\omega$ (linear in $\omega$ means there exist constants $c_0, c_1,$ with $c_1 \in \mathbb{Z},$ such that, $\arg T(e^{j\omega}) = c_1 \omega + c_0$), we say that the output $X_{\text{out}}(e^{j\omega})$ of the filter bank suffers from phase distortion ($\arg T(e^{j\omega})$).

**Definition 7.2.1** If a filter bank is free from aliasing, amplitude distortion, and phase distortion, it is said to have the perfect reconstruction property. This means there exist a constant $c$ and an integer $n_0$ in $\mathbb{N}_0$ such that:

$$T(z) = cz^{-n_0}.$$  

This definition ignores the coding/decoding error and filter roundoff noise. Roundoff noise or quantization errors are a consequence of coefficient quantization and signal quantization. The coding/decoding errors are a consequence of the compression and the decompression process in the channel. In fact, signal quantization can be considered as a coding/decoding error, for more details we refer to [89]. Now we summarize the perfect reconstruction property in the following: If the analysis and the synthesis filters, $A_m(z)$ and $S_m(z), m = 0, \ldots, M - 1,$ are such that:

- Aliasing is completely canceled, and
- $T(z)$ is a pure delay, that is, $T(z) = cz^{-t_0}, c \neq 0,$

then the filter bank is free from aliasing, amplitude distortion and phase distortion. The system satisfies $x_{\text{out}}(t) = cx_{\text{in}}(t - t_0),$ which means perfect reconstruction.
In terms of the polyphase matrices $E(z)$ and $R(z)$, see (7.4) and (7.6), a sufficient condition for perfect reconstruction is given by:

$$R(z)E(z) = cz^{-\tau_0}I,$$

for some integer $\tau_0$ and a constant $c \neq 0$. If this condition holds, then we have

$$T(z) = cz^{L\tau_0 + M - 1}.$$

The necessary and sufficient condition for perfect reconstruction is given in the following theorem and proved in [89, 90].

**Theorem 7.2.2** A filter bank has perfect reconstruction if and only if

$$R(z)E(z) = cz^{-\tau_0} \begin{bmatrix} 0 & I_{L-t} \\ z^{-1}I_t & 0 \end{bmatrix},$$

for some integer $0 \leq l \leq L - 1$, integer $\tau_0$, constant $c \neq 0$. □

**Proof:** See Subsection 5.6.2 in [89]. ■

In the oversampled case, the reconstruction scheme is not uniquely determined. This non-uniqueness entails a desirable freedom that can be used for design. The following theorem presents a parameterization of all perfect reconstructing synthesis polyphase matrices, given an analysis polyphase matrix.

**Theorem 7.2.3** [15] Assuming that the analysis polyphase matrix $E(z)$, $(M > L)$, has full rank almost everywhere, any synthesis polyphase matrix $R(z)$ providing perfect reconstruction can be written as

$$R(z) = R_0(z) + U(z) [I - E(z)R_0(z)],$$

where $R_0(z)$ is any particular perfect reconstructing synthesis polyphase matrix, and $U(z)$ any $L \times M$ matrix, satisfying $|U(e^{2\pi l})| < \infty, l = 0, \ldots, L - 1, m = 0, \ldots, M - 1$. A special choice of $R_0(z)$ is the pseudo-inverse of $E(z)$:

$$R_0(z) = (E^*(1/z^*)E(z))^{-1}E^*(1/z^*).$$

□

**Proof:** See p. 113 in [13]. ■

Of course, if the analysis polyphase matrix $E(z)$ is square and invertible, then the inverse is unique.
7.3 Equivalence between filter banks and frames

We start from the analysis filter bank as shown in Fig. 7.4, where \((A_m(x))_{m=0, \ldots, M-1}\) are the analysis filters. Let \(a_m, m = 0, \ldots, M - 1\) be the corresponding impulse responses of \(A_m(x), m = 0, \ldots, M - 1\), then we have

\[
y_{\text{out}}(l) = (L(a_m \ast x_{\text{in}}))(l) = \sum_{t \in \mathbb{Z}} x_{\text{in}}(t)a_m(l - t), \quad m = 0, \ldots, M - 1,
\]

for all \(l \in \mathbb{Z}\), where \(\sigma\) is the right-shift operator:

\[
(\sigma x)(t) = x(t - 1).
\]

We shall show that the input signal \(x_{\text{in}}\) can be reconstructed from \(y_{\text{out}}\) if the system \(\{\sigma l a^*_m\}_{l \in \mathbb{Z}, m = 0, \ldots, M - 1}\) constitutes a frame in \(\ell_2(\mathbb{Z})\). A frame in \(\ell_2(\mathbb{Z})\) is an ‘over’-complete system in \(\ell_2(\mathbb{Z})\) that satisfies a stability condition in \(\ell_2(\mathbb{Z})\). Specifically for the system \(\{\phi_n | n \in \Gamma\}\) to be a frame in \(\ell_2(\mathbb{Z})\), where \(\Gamma\) is an arbitrary countable set, we have the following definition.

**Definition 7.3.1** The the system \(\{\phi_n | n \in \Gamma\}\) in \(\ell_2(\mathbb{Z})\) is said to be a frame in \(\ell_2(\mathbb{Z})\) if there exist \(A > 0, B > 0\) such that:

\[
\forall x \in \ell_2(\mathbb{Z}), \quad A\|x\|^2 \leq \sum_{n \in \Gamma} |\langle x, \phi_n \rangle|^2 \leq B\|x\|^2. \tag{7.16}
\]

The optimal \(A\) and \(B\) are called the frame bounds. The frame \(\{\phi_n | n \in \Gamma\}\) is said to be tight if \(A = B\). A frame is exact if it is no longer a frame whenever any one of its elements is removed. Otherwise the frame is called overcomplete.

For a frame \(\{\phi_n | n \in \Gamma\}\), the series \(\sum_{n \in \Gamma} |\langle x, \phi_n \rangle|^2\) converges, and obviously frames are complete systems. A system \(\{\phi_n | n \in \Gamma\}\) in the Hilbert space \(\ell_2(\mathbb{Z})\) is an exact frame if and only if it is a Riesz\(^1\) basis in \(\ell_2(\mathbb{Z})\), as can be found in [9, 102].

Let \(\{\phi_n | n \in \Gamma\}, \Gamma \subseteq \mathbb{Z}^d, d \in \mathbb{N}\), be a frame in \(\ell_2(\mathbb{Z})\), and define the frame operator \(A\) on \(\ell_2(\mathbb{Z})\) by:

\[
Ax = \sum_{n \in \Gamma} \langle x, \phi_n \rangle e_n,
\]

\(^1\)A countable system \(\{\phi_n | n \in \Gamma\}\) is said to be a Riesz basis in \(\ell_2(\mathbb{Z})\) if there exists a bounded invertible operator \(V \in \mathcal{B}(\ell_2(\Gamma), \ell_2(\mathbb{Z}))\) such that \(Ve_n = \phi_n\), with \(\{e_n | n \in \Gamma\}\) the standard orthonormal basis in \(\ell_2(\Gamma)\). In this thesis, we make no difference in using the names Riesz basis and exact frame.
where \( \{e_n\mid n \in \Gamma\} \) is the standard orthonormal basis in \( \ell_2(\mathbb{Z}) \), \((A)\phi_n = \langle x, \phi_\cdot \rangle \).

For the frame operator, we have the following immediate result if the system \( \{\phi_n\mid n \in \Gamma\} \) is a frame in \( \ell_2(\mathbb{Z}) \). From the Definition 7.3.1 of frames, the condition (7.16) can be reformulated as follows:

\[
\forall x \in \ell_2(\mathbb{Z}), \quad A\|x\|^2 \leq \|Ax\|^2 \leq B\|x\|^2,
\]

so the following lemma.

**Lemma 7.3.2** Let \( \{\phi_n\mid n \in \Gamma\} \) be a frame in \( \ell_2(\mathbb{Z}) \). Then the frame operator \( A \) is a well-defined, bounded operator, which is bounded from below, from \( \ell_2(\mathbb{Z}) \) in \( \ell_2(\Gamma) \).

**Definition 7.3.3** Let \( \{\phi_n\mid n \in \Gamma\} \) be a frame in \( \ell_2(\mathbb{Z}) \). The analysis operator \( J \) associated with the frame \( \{\phi_n\mid n \in \Gamma\} \) is defined by

\[
J = A^* A
\]

where \( A \) is the frame operator.

**Lemma 7.3.4** The system \( \{\phi_n\mid n \in \Gamma\} \) in a Hilbert space \( \ell_2(\mathbb{Z}) \) is a frame with frame bounds \( A, B \) if and only if the associated analysis operator \( J \) is a bounded linear operator with

\[
AI \leq J \leq BI.
\]

**Proof:** Let \( \{\phi_n\mid n \in \Gamma\} \) be a frame in \( \ell_2(\mathbb{Z}) \). Computing \( J = A^* A \), we have the explicit form of the analysis operator \( J \):

\[
Jx = \sum_{n \in \Gamma} \langle x, \phi_n \rangle \phi_n
\]

for all \( x \in \ell_2(\mathbb{Z}) \). Consequently, we have:

\[
\langle Jx, x \rangle = \sum_{n \in \Gamma} |\langle x, \phi_n \rangle|^2,
\]

this can be also derived from \( J = A^* A \).

From the definition of frames there exist \( A > 0, B > 0 \) such that:

\[
\forall x \in \ell_2(\mathbb{Z}), \quad A\|x\|^2 \leq \sum_{n \in \Gamma} |\langle x, \phi_n \rangle|^2 \leq B\|x\|^2.
\]
which is equivalent to

\[
\forall x \in \ell_2(\mathbb{Z}), \quad A\|x\|^2 \leq \langle Jx, x \rangle \leq B\|x\|^2,
\]

or to

\[
AI \leq J \leq BI.
\]

The frame bounds can be determined by:

\[
A = \inf\{\|Ax\|^2 | x \in \ell_2(\mathbb{Z}), \|x\| = 1\},
\]

\[
B = \sup\{\|Ax\|^2 | x \in \ell_2(\mathbb{Z}), \|x\| = 1\}.
\]

The injectivity of the frame operator \( A \) is equivalent with completeness of the system \( \{\phi_n | n \in \Gamma\} \), that is, \( \{\phi_n | n \in \Gamma\}^\perp = \{0\} \). Boundedness of the operator \( A \) implies that a small perturbation \( \delta x \) of a sequence \( x \) in \( \ell_2(\mathbb{Z}) \) results in a small perturbation in \( \mathcal{A}x \) in \( \ell_2(\Gamma) \), and finally, boundedness below of \( A \) implies that for \( x \in \ell_2(\mathbb{Z}) \) significantly not equal to \( 0 \), that is, \( \|x\|^2 \geq \varepsilon > 0 \), the sequence \( \mathcal{A}x \) is significantly not equal to the null sequence in \( \ell_2(\Gamma) \). Otherwise said, the correspondence \( x \mapsto \mathcal{A}x \) is well-conditioned.

Any bounded left inverse \( B \in \mathcal{B}(\ell_2(\Gamma), \ell_2(\mathbb{Z})) \) of \( A \) yields the possibility to obtain \( x \) from \( \mathcal{A}x \), namely \( x = B(\mathcal{A}x) \) and so

\[
x = \sum_{n \in \Gamma} \langle x, \phi_n \rangle B e_n.
\]

Since \( B \) is surjective, the collection \( \{Be_n | n \in \Gamma\} \) is a frame with frame operator \( B^* \). The frame \( \{Be_n | n \in \Gamma\} \) is called a dual frame of the frame \( \{\phi_n | n \in \Gamma\} \). However, if the computation is not exact, instead of \( \mathcal{A}x \) we obtain \( \mathcal{A}x + \varepsilon x \) with \( \varepsilon x \in \ell_2(\Gamma) \). And then applying \( B \) yields

\[
B(\mathcal{A}x + \varepsilon x) = x + B\varepsilon x.
\]

So we want to minimize \( \|B\varepsilon x\| \) for all \( B \in \mathcal{B}(\ell_2(\Gamma), \ell_2(\mathbb{Z})) \) with \( B\mathcal{A} = \mathcal{I} \). In the following proposition, \( \|B\| \) denotes the norm of the operator \( B \), which is defined by

\[
\|B\| := \sup \{ \|By\| | y \in \ell_2(\Gamma), \|y\| = 1 \}.
\]

**Proposition 7.3.5** Let \( A \in \mathcal{B}(\ell_2(\mathbb{Z}), \ell_2(\Gamma)) \) be bounded from below. Then the operator \((A^* A)^{-1} A^*\) has the property

\[
\| (A^* A)^{-1} A^* \| = \min \{ \|B\||B\mathcal{A} = \mathcal{I}, B \in \mathcal{B}(\ell_2(\Gamma), \ell_2(\mathbb{Z})) \}
\]

\( \square \)
7.3 Equivalence between filter banks and frames

**Proof:** Define $\mathcal{P} := A (A^* A)^{-1} A^*$. It is easily verified that $\mathcal{P}^2 = \mathcal{P} = \mathcal{P}^*$, so that $\mathcal{P}$ is an orthogonal projection. Moreover $\mathcal{P} A = A$ and $\mathcal{P} = A (A^* A)^{-1} A^*$ implying that $\text{range}(\mathcal{P}) = \text{range}(A)$. So $\mathcal{P}$ is the orthogonal projection of $\ell_2(\Gamma)$ onto $\text{range}(A)$.

Write now $B_0 = (A^* A)^{-1} A^*$ and let $B \in \mathcal{B}(\ell_2(\Gamma), \ell_2(\mathbb{Z}))$ satisfy $B A = \mathcal{I}$. Then $B_0 \mathcal{P} = B_0$ and $B \mathcal{P} = B_0$. Hence

$$\|B_0\| = \|B_0 \mathcal{P}\| = \|B \mathcal{P}\| \leq \|B\|.$$  

$\blacksquare$

**Theorem 7.3.6** [9, 102] Let the system $\{\phi_n | n \in \Gamma\}$ be a frame in a Hilbert space $\ell_2(\mathbb{Z})$. Then $\mathcal{J}$ is a self-adjoint invertible operator and $B^{-1} \mathcal{I} \leq \mathcal{J}^{-1} \leq A^{-1} \mathcal{I}$. Further, for all $x \in \ell_2(\mathbb{Z})$,

$$x = \sum_{n \in \Gamma} \langle x, \mathcal{J}^{-1} \phi_n \rangle \phi_n$$

and

$$x = \sum_{n \in \Gamma} \langle x, \phi_n \rangle \mathcal{J}^{-1} \phi_n.$$ 

The system $\{\mathcal{J}^{-1} \phi_n | n \in \Gamma\}$ is a frame in $\ell_2(\mathbb{Z})$, with frame bounds $1/B$ and $1/A$. $\{\mathcal{J}^{-1} \phi_n | n \in \Gamma\}$ is called the dual frame of the frame $\{\phi_n | n \in \Gamma\}$. $\Box$

**Proof:** Follows directly from Lemma 7.3.4, $\frac{1}{A} \mathcal{J}^{-1} (\mathcal{J} - A \mathcal{I}) \geq 0$. $\blacksquare$

Since the analysis operator $\mathcal{J}$ is self-adjoint and positive we have for the frame bounds:

$$A = \|\mathcal{J}^{-1}\|^{-1},$$

$$B = \|\mathcal{J}\| = \|A\|^2.$$ 

**Theorem 7.3.7** [9, 102] Let the system $\{\phi_n | n \in \Gamma\}$ be a frame in the Hilbert space $\ell_2(\mathbb{Z})$. Then $\{\phi_n | n \in \Gamma\}$ is a Riesz basis in $\ell_2(\mathbb{Z})$ if and only if $A$ is invertible. In this case $\{\phi_n | n \in \Gamma\}$ and $\{\mathcal{J}^{-1} \phi_n | n \in \Gamma\}$ are bi-orthonormal, that is,

$$\langle \mathcal{J}^{-1} \phi_m, \phi_n \rangle = \delta_{mn}.$$ 

for all $m, n \in \Gamma$. The system $\{\mathcal{J}^{-1} \phi_n | n \in \Gamma\}$ is the unique system which is bi-orthonormal to the system $\{\phi_n | n \in \Gamma\}$. $\Box$
Proof: See Theorem 3.5 in [9].

For the reconstruction of \( x \) from a given \( y = \mathcal{J}x \), tight frames are of particular importance, since their dual frame is easier to determine. Indeed, let \( \{\phi_n|n \in \Gamma\} \) be a frame in \( \ell_2(\mathbb{Z}) \) with frame operator \( A \), and frame bounds \( A, B \), with \( B = A \), then we see that \( \mathcal{J} = A \mathcal{I} \). So for the dual frame \( \{\psi_n|n \in \Gamma\} \) we have

\[
\psi_n = (A^* A)^{-1} \phi_n = \frac{1}{A} \phi_n, \quad n \in \Gamma.
\]

If \( A = 1 \), then the dual frame is the frame itself. In this case it is also called a self-dual frame. For each \( x \in \ell_2(\mathbb{Z}) \) we have

\[
x = \sum_{n \in \Gamma} \langle x, \phi_n \rangle \phi_n.
\]

A self-dual frame is not necessarily orthogonal. This can be seen as follows:

\[
\sum_{n \in \Gamma \setminus \{m\}} |\langle \phi_m, \phi_n \rangle|^2 = \|\phi_m\|^2(1 - \|\phi_m\|^2).
\]

Consequently, a self-dual frame \( \{\phi_n|n \in \Gamma\} \), with \( \|\phi_n\| = 1 \) for all \( n \in \Gamma \), is a complete orthonormal system.

We return to the representation of the analysis filter bank (7.15): suppose that the system \( \{\sigma^L \tilde{a}_m^*|l \in \mathbb{Z}, m = 0, \ldots, M - 1\} \) is a frame in \( \ell_2(\mathbb{Z}) \), the analysis operator \( \mathcal{J} \) associated to \( \{\sigma^L \tilde{a}_m^*|l \in \mathbb{Z}, m = 0, \ldots, M - 1\} \) then reads:

\[
\mathcal{J} x_{\text{in}} = \sum_{m=0}^{M-1} \sum_{l \in \mathbb{Z}} \langle x_{\text{in}}, \sigma^L \tilde{a}_m^* \rangle \sigma^L \tilde{a}_m^* = \sum_{m=0}^{M-1} \sum_{l \in \mathbb{Z}} y_{\text{out}}(l) \sigma^L \tilde{a}_m^*.
\]  

(7.17)

Now we can reconstruct a signal \( x_{\text{in}} \) from \( y_{\text{out}} \) by:

\[
x_{\text{in}} = \sum_{m=0}^{M-1} \sum_{l \in \mathbb{Z}} y_{\text{out}}(l) \mathcal{J}^{-1} \sigma^L \tilde{a}_m^*.
\]  

(7.18)

Lemma 7.3.8 The analysis operator \( \mathcal{J} \) commutes with \( \sigma^L \), for all \( l \in \mathbb{Z} \), that is,

\[
\sigma^L \mathcal{J} = \mathcal{J} \sigma^L.
\]

\[\square\]
7.3 Equivalence between filter banks and frames

**Proof:** It is sufficient to prove $\sigma^I \mathcal{J} = \mathcal{J} \sigma^I$:

$$
\mathcal{J} \sigma^I x = \sum_{m=0}^{M-1} \sum_{l \in \mathbb{Z}} \langle \sigma^I x, \sigma^I \hat{a}_m^* \sigma^I \hat{a}_m^* \rangle 
= \sum_{m=0}^{M-1} \sum_{l \in \mathbb{Z}} \langle x, \sigma^{(l-1)} \hat{a}_m^* \sigma^{(l-1)} \hat{a}_m^* \rangle 
= \sum_{m=0}^{M-1} \sum_{l \in \mathbb{Z}} \langle x, \sigma^{(l+1)} \hat{a}_m^* \sigma^{(l+1)} \hat{a}_m^* \rangle 
= \sigma^I \sum_{m=0}^{M-1} \sum_{l \in \mathbb{Z}} \langle x, \sigma^{(l+1)} \hat{a}_m^* \sigma^{(l+1)} \hat{a}_m^* \rangle.
$$

Consequently, for the reconstruction of $x_{in}$ from $y_{out}$ we have now:

$$
x_{in} = \sum_{m=0}^{M-1} \sum_{l \in \mathbb{Z}} y_{out}(l) \sigma^I \mathcal{J}^{-1} \hat{a}_m^*. \quad (7.19)
$$

From (7.19), the reconstruction scheme using filter banks can be realized with a synthesis filter bank as shown in Fig. 7.5. Here the impulse response $s_m$ of $S_m$, $m = 0, \ldots, M - 1$, satisfies

$$
s_m = \mathcal{J}^{-1} \hat{a}_m^*, \quad m = 0, \ldots, M - 1. \quad (7.20)
$$

The synthesis filter bank is stable (perhaps not BIBO stable) in the sense that $\mathcal{J}^{-1}$ is bounded from below in $\mathcal{B}(\ell_2(\mathbb{Z}))$, see Theorem 7.3.6.

If the frame $\{ \phi_n | n \in \Gamma \}$ is not tight, the computation of $\mathcal{J}^{-1}$, and therefore its dual frame, is not easy. Of course, in the case of finite-dimensional signal spaces the frame bounds can be determined by straightforward matrix calculus. In this case we can use a truncation of Neumann series to approximate $\mathcal{J}^{-1}$. Let $\{ \phi_n | n \in \Gamma \}$ be a frame in $\ell_2(\mathbb{Z})$, and $\mathcal{J}$ the associated analysis operator, and consider the operator $\mathcal{I} - \frac{2}{B + A} \mathcal{J}$. From (7.3.4), we have

$$
-B - A \mathcal{I} \leq \mathcal{I} - \frac{2}{B + A} \mathcal{J} \leq \frac{B - A}{B + A} \mathcal{I}.
$$

Thus $\| \mathcal{I} - \frac{2}{B + A} \mathcal{J} \| \leq \frac{B - A}{B + A} < 1$, and the Neumann series $\sum_{k=0}^{\infty} (\mathcal{I} - \frac{2}{B + A} \mathcal{J})^k$ converges in $\mathcal{B}(\ell_2(\mathbb{Z}))$ to $\frac{B + A}{B + A} \mathcal{J}^{-1}$. Hence

$$
\mathcal{J}^{-1} = \frac{2}{B + A} \sum_{k=0}^{\infty} (\mathcal{I} - \frac{2}{B + A} \mathcal{J})^k.
$$
The rate of convergence depends on how small \((B-A)/(B+A)\) is. If \((B-A)/(B+A) \ll 1\), in which case, one says the frame is ‘snugged’, we truncate the Neumann series and write:

\[
\mathcal{J}^{-1} \approx \frac{2}{A+B} \sum_{k=0}^{K} (\mathcal{I} - \frac{2}{A+B} \mathcal{J})^k.
\]

For the dual frame \(\{\psi_n | n \in \Gamma\}\) we have:

\[
\psi_n = \frac{2}{A+B} \sum_{k=0}^{\infty} (\mathcal{I} - \frac{2}{A+B} \mathcal{J})^k \phi_n, \ n \in \Gamma.
\]

If we approximate \(\psi_n\) by \(\frac{2}{A+B} \sum_{k=0}^{K} (\mathcal{I} - \frac{2}{A+B} \mathcal{J})^k \phi_n\), we have for the truncation error:

\[
||\psi_n - \frac{2}{A+B} \sum_{k=0}^{K} (\mathcal{I} - \frac{2}{A+B} \mathcal{J})^k \phi_n|| = \frac{2}{A+B} \sum_{k=0}^{\infty} ||\mathcal{I} - \frac{2}{A+B} \mathcal{J}||^k ||\phi_n||
\]

\[
\leq \frac{2}{A+B} \sum_{K+1}^{\infty} ||\mathcal{I} - \frac{2}{A+B} \mathcal{J}||^k ||\phi_n||
\]

\[
= \frac{2}{A+B} \left(1 - \frac{2}{A+B} \mathcal{J}|| \phi_n||
\right)
\]

\[
\leq \frac{2}{A+B} \left(1 - \frac{2}{A+B} \mathcal{J}|| \phi_n||
\right)
\]

\[
= \frac{1}{A} \left(\frac{B-A}{B+A} \right)^{K+1} ||\phi_n||.
\]

A good approximation of \(\mathcal{J}^{-1}\) is obtained if \(A \approx B\), by taking one or two terms of the Neumann series. In many practical situations, the true values of the frame bounds are not known or hard to calculate. On the other hand, insufficient information about the frame bounds may lead to a poor rate of convergence. In these cases we can replace \(2/(A+B)\) in the Neumann series by a small enough \(\gamma\), see [35]:

\[
\mathcal{J}^{-1} = \gamma \sum_{k=0}^{\infty} (\mathcal{I} - \gamma \mathcal{J})^k.
\]

The computation of \(\mathcal{J}^{-1}x\) can be iteratively obtained by the frame algorithm:

\[
x_0 = \gamma x,
\]

\[
x_{k+1} = (\mathcal{I} - \gamma \mathcal{J})x_k + \gamma x.
\]
7.3 Equivalence between filter banks and frames

In [35] it is shown that the optimal choice of $\gamma$ is $2/(B + A)$. We conclude this discussion by saying that computing the dual frame by the computation of $J^{-1}$ is quite expensive, especially if $(B - A)/(B + A)$ is not much below 1. Furthermore, the dual frame elements $s_m$, $m = 0, \ldots, M - 1$, have to be impulse responses of filters, that is, the $z$-transform of $s_m$ in $C(z)$.

In the following, we translate the problem of computing the frame bounds to the $z$-domain. We define the $M \times L$ matrix-valued function $H(z) = H(z)$ such that

$$H(z) = A_m(\Omega_L z)$$

with $m = 0, \ldots, M - 1$, $l = 0, \ldots, L - 1$, and $\Omega_L = e^{-j \frac{2\pi}{L}}$. Note that $H(z)$ is (the transpose of) the alias component matrix obtained from the analysis filter bank with filters $a_m$, $m = 0, \ldots, M - 1$. Define the $L \times L$ matrix-valued function $G(z) = H^*(z)H(z)$, for all $z \in \mathbb{U}$ such that $0 \leq \arg(z) < \frac{2\pi}{L}$. For all $z \in \mathbb{U}$, the matrix $G(z)$ is invertible if and only if $H(z)$ is injective.

**Theorem 7.3.9** [92] The system $\{\sigma^L \hat{a}_m^*| l \in \mathbb{Z}, m = 0, \ldots, M - 1\}$ is a frame in $\ell_2(\mathbb{Z})$, with bounds $A$ and $B$, if and only if

$$AI \leq H^*(z)H(z) \leq BI$$

for all $z \in \mathbb{U}$, $0 \leq \arg(z) < \frac{2\pi}{L}$. Consequently $L \leq M$.

**Proof:** See Chapter 6 in [92].

A necessary condition such that $\{\sigma^L \hat{a}_m^*| l \in \mathbb{Z}, m = 0, \ldots, M - 1\}$ is a frame in $\ell_2(\mathbb{Z})$ is given by $L \leq M$. In the special case $L = M$ we get a Riesz basis in $\ell_2(\mathbb{Z})$.

Analysis filter banks can be seen as frame operators. Their properties with respect to frames are identical:

$$AI \leq A^* A \leq BI \iff AI \leq H^*(z)H(z) \leq BI,$$

which is also equivalent to $AI \leq E^*(1/z^*)E(z) \leq BI$, on $\mathbb{U}$, see also [15, 92], where $E(z)$ is the analysis polyphase matrix; see Section 7.2. We note that $E^*(1/z^*) = E^*(z)$, on $\mathbb{U}$. Furthermore, the frame bounds are in terms of the extremes of the eigenvalues of $H^*(z)H(z)$, on $\mathbb{U}$. A method to compute these values is described in [35, 49, 92].

**Theorem 7.3.10** [15, 34, 35, 49, 92] The frame bounds $A$ and $B$ of a frame providing a filter bank are given by the essential infimum and supremum, respectively, of the eigenvalues $\lambda_i(\theta)$ of the matrix $E^*(e^{j2\pi \theta})E(e^{j2\pi \theta})$:

$$A = \text{ess inf}_{\theta \in [0,1]} \lambda_i(\theta),$$

$$B = \text{ess sup}_{\theta \in [0,1]} \lambda_i(\theta).$$


Proof: See [15, 92].

In the following theorem we give the necessary and sufficient condition for a system, obtained by an analysis bank, to be a frame.

**Theorem 7.3.11** [15, 33] The system \( \{ \sigma^l a_m^* \}, l \in \mathbb{Z}, m = 0, \ldots, M - 1 \) is a frame in \( \ell_2(\mathbb{Z}) \) if and only if the analysis polyphase matrix has a full rank on the unit circle:

\[
\text{rank} (\mathbf{E}(z)) = L, \ z \in \mathbb{U}
\]

Proof: See Theorem 5.1 in [15].

The information output \( \{ y_{\text{outm}}(l) \}, l \in \mathbb{Z}, m = 0, \ldots, M - 1 \) of the analysis filter bank can be seen as a ‘sub’-sampling of a localized transform. In other words, the system \( \{ \sigma^l a_m^* \}, l \in \mathbb{Z}, m = 0, \ldots, M - 1 \) can be seen as a subsampled version of a kernel of a localizing transformation. This is already the case for the DFT filter banks, where the system \( \{ \sigma^l a_m^* \}, l \in \mathbb{Z}, m = 0, \ldots, M - 1 \) corresponds with a subsampling of the kernel of the short-time Fourier transformation, for more details see literature [89]. In the next section, we consider a more general case of Causal Sliding Windowed Unitary Transformations (CSWUT), where the kernel is not necessarily the one of the short-time Fourier transformation. In Section 7.6, we give illustrations of the CSWUT, using representations from Part I.

### 7.4 Localizing unitary transformations and filter banks

In order to define a localizing unitary transformation, we introduce a kernel-function \( B(t, \omega) \) for \( t \in \mathbb{Z} \) and \( \omega \in \mathcal{J} \), where \( \mathcal{J} \) is a subset of \( \mathbb{R} \). The subset \( \mathcal{J} \) will be an interval or a discrete subset in \( \mathbb{R} \). The kernel-function \( B(t, \omega) \) is a bounded two-variable function, with a discrete-variable \( t \) and a ‘continuous’-variable \( \omega \), satisfying the so-called bi-orthogonality conditions:

\[
\int_{\mathcal{J}} B(t, \omega) B^*(t', \omega) d\omega = \delta_{tt'}, \quad (7.22)
\]

\[
\sum_{t \in \mathbb{Z}} B(t, \omega) B^*(t, \omega') = \delta(\omega - \omega'), \quad (7.23)
\]

where \( \delta(\omega) \) is the Dirac ‘function’, to be interpreted as \( \delta(\omega - \omega') = \delta_{\omega\omega'} \) if \( \mathcal{J} \) is a discrete set, and the integral \( \int_{\mathcal{J}} \) becomes \( \sum_{\mathcal{J}} \) in this case, where \( \delta_{\mathcal{J}t'} \) is the Kronecker delta:

\[
\delta_{tt'} = \begin{cases} 
1 & \text{if } t = t', \\
0 & \text{otherwise.}
\end{cases}
\]
7.4 Localizing unitary transformations and filter banks

An example of a kernel function is: $B(t, \omega) = \frac{1}{\sqrt{2\pi}} e^{it\omega}$, $t \in \mathbb{Z}$, $\omega \in \mathbb{J} = [-\pi, \pi)$, the kernel of the discrete-time Fourier transformation. Here the convergence is taken in the distributional sense. Such a kernel function can be constructed as follows: Let the two systems $\{\phi_n | n \in \mathbb{N}_0\}$ and $\{\psi_n | n \in \mathbb{N}_0\}$ be orthonormal bases in $\ell_2(\mathbb{Z})$ and $L_2(\mathbb{J})$, respectively. Then we have:

$$\sum_{n \in \mathbb{N}_0} \phi_n(t)\phi_n(t') = \delta_{tt'},$$  \hspace{1cm} (7.24)

$$\sum_{n \in \mathbb{N}_0} \psi_n(\omega)\psi_n(\omega') = \delta(\omega - \omega').$$  \hspace{1cm} (7.25)

For the kernel function $B(t, \omega)$ we write:

$$B(t, \omega) = \sum_{n \in \mathbb{N}_0} \phi_n^*(t)\psi_n(\omega), \hspace{1cm} (t, \omega) \in \mathbb{Z} \times \mathbb{J},$$  \hspace{1cm} (7.26)

assuming that $\sum_{n \in \mathbb{N}_0} \phi_n^*(t)\psi_n(\omega)$ converges pointwise for $(t, \omega) \in \mathbb{Z} \times \mathbb{J}$. Then the kernel function $B(t, \omega)$ satisfies (7.22) and (7.23). An example of this situation occurs if we choose for $\{\phi_n | n \in \mathbb{N}_0\}$ the Laguerre or Kautz basis, and for $\{\psi_n | n \in \mathbb{N}_0\}$ the standard orthonormal basis; see Section 7.6. The kernel function $B(t, \omega)$ is identified as the kernel of the unitary transformation $\mathcal{U}$ on $\ell_2(\mathbb{Z})$

$$(\mathcal{U}x)(\omega) = \sum_{t \in \mathbb{Z}} x(t)B(t, \omega), \hspace{1cm} \omega \in \mathbb{J},$$  \hspace{1cm} (7.27)

and $\mathcal{U}$ is an isometry from $\ell_2(\mathbb{Z})$ into $L_2(\mathbb{J})$ [1, 101]. Then the adjoint $\mathcal{U}^*$ of $\mathcal{U}$ on $\ell_2(\mathbb{Z})$ satisfies

$$(\mathcal{U}^*y)(t) = \sum_{n \in \mathbb{N}_0} \langle x, \psi_n \rangle \phi_n(t) = \int_{\mathbb{J}} y(\omega)B^*(t, \omega)d\omega,$$

The choice of such a kernel $B(t, \omega)$ leads to the definition of the (amplitude) spectrum $\mathcal{U}x$. The spectrum $(\mathcal{U}x)(\omega)$ of $x$ can be interpreted as a global distribution of the energy of the signal $x$ as function of $\omega$. Knowing the spectrum $(\mathcal{U}x)(\omega)$ of $x$ for all $\omega \in \mathbb{J}$, we can determine $x(t)$ for all $t \in \mathbb{Z}$. Of course, the concepts are known for $B(t, \omega) = \frac{1}{\sqrt{2\pi}} e^{it\omega}$, $t \in \mathbb{Z}$, $\omega \in \mathbb{J} = [-\pi, \pi)$, where the resulting $(\mathcal{U}x)(\omega)$ is the Fourier spectrum of the signal $x$. The kernel function $B(t, \omega) = \frac{1}{\sqrt{2\pi}} e^{it\omega}$, can be constructed by (7.26), if we take $\phi_n(t) = \delta_n(t)$ and $\psi_n(\omega) = \frac{1}{\sqrt{2\pi}} e^{in\omega}$, where $\delta_n(t) = \delta_{nt}$.

However, one is often more interested in the momentary or local distribution of the energy as a function of $\omega$. A generally accepted idea to localize the spectrum
Unif ormly downsampled filter banks

\( \mathcal{U}x \) is the use of a window function \( w \in \ell_2(\mathbb{Z}) \). We define the set of all admissible anti-causal windows by \( \mathbb{W}_c \), that is, windows satisfying:

\[
w \in \mathbb{W}_c :\iff w \in \ell_1(\mathbb{Z}) \setminus \{0\}, \text{ real-valued, and } \text{support}(w) \subseteq \mathbb{Z}_0^-.\]

**Definition 7.4.1** Let \( B(t, \omega) \) be the kernel function of a unitary transformation, and \( w \) an admissible window. The (discrete-time) Causal Sliding Windowed Unitary Transformation (CSWUT), denoted by \( \mathcal{W}_w \), or shortly \( \mathcal{W} \), is defined by:

\[
(\mathcal{W}x)(a, \omega) = \langle \mathcal{U}(w \cdot \sigma^{-a}x) \rangle(\omega) = \sum_{t \in \mathbb{Z}} x(t)w(t-a)B(t-a, \omega),
\]

for \( (a, \omega) \in \mathbb{Z} \times \mathbb{J}, x \in \ell_2(\mathbb{Z}) \), where

\[
k(t, \omega) = w(t)B^*(t, \omega).
\]

\( w \cdot \sigma^{-a}x \) denotes the pointwise product of \( w \) and \( \sigma^{-a}x \).

The kernel function function \( k(t, \omega) \) of the CSWUT satisfies \( k(\cdot, \omega) \in \ell_1(\mathbb{Z}) \), since \( w \in \mathbb{W}_c \) and \( B(t, \omega) \) is bounded on \( \mathbb{Z} \times \mathbb{J} \). By the Cauchy-Schwarz inequality we have \( |(\mathcal{W}x)(a, \omega)| \leq \|x\\| k(\cdot, \omega) \|, \) for \( x \in \ell_2(\mathbb{Z}) \).

**Theorem 7.4.2** For all \( x_1, x_2 \in \ell_2(\mathbb{Z}) \) and admissible windows \( w_1, w_2 \in \mathbb{W}_c \)

\[
\langle (\mathcal{W}_{w_1}x_1)(a, \cdot), (\mathcal{W}_{w_2}x_2)(a, \cdot) \rangle = \langle x_1 \cdot \sigma^a w_1, x_2 \cdot \sigma^a w_2 \rangle, a \in \mathbb{Z}.
\]

\[
\langle \mathcal{W}_{w_1}x_1, \mathcal{W}_{w_2}x_2 \rangle = \langle w_1, w_2 \rangle \langle x_1, x_2 \rangle.
\]

**Proof:** From the definition of the CSWUS, see Definition 7.4.1, we have

\[
(\mathcal{W}_{w}x)(a, \omega) = (\mathcal{U}w \cdot \sigma^{-a}x)(\omega),
\]

where \( \mathcal{U} \) is unitary operator. Thus,

\[
\langle (\mathcal{W}_{w_1}x_1)(a, \cdot), (\mathcal{W}_{w_2}x_2)(a, \cdot) \rangle = \langle \mathcal{U}(w_1 \cdot \sigma^{-a}x_1), \mathcal{U}(w_2 \cdot \sigma^{-a}x_2) \rangle,
\]

\[
= \langle w_1 \cdot \sigma^{-a}x_1, w_2 \cdot \sigma^{-a}x_2 \rangle,
\]

\[
= \langle \sigma^a w_1 \cdot x_1, \sigma^a w_2 \cdot x_2 \rangle.
\]

Taking the summation over \( a \) in (7.28) leads to

\[
\langle (\mathcal{W}_{w_1}x_1), (\mathcal{W}_{w_2}x_2) \rangle = \langle w_1, w_2 \rangle \langle x_1, x_2 \rangle.
\]
Given a window function \( w \in \mathcal{W}_c \), a signal sequence \( x \in \ell_2(\mathbb{Z}) \) is uniquely determined by its Causal Sliding Windowed Unitary Spectrum (CSWUS) \( \mathcal{W}x \). The CSWUS is an intermediate signal description between the pure \( t \) and the pure \( \omega \) representation. In the following lemma we give a reconstruction formula of a signal \( x \) from its CSWUS.

**Theorem 7.4.3** Given a CSWUS \( g = \mathcal{W}x \), then the signal \( x \) can reconstructed from \( g \) by:

\[
x(t) = \frac{1}{\| w \|^2} \sum_{a \in \mathbb{Z}} \int_{\mathbb{R}} g(t - a, \omega)B^*(a, \omega)d\omega, \quad t \in \mathbb{Z}.
\] (7.29)

Furthermore, the CSWUS \( \mathcal{W}x \) minimizes \( \| g \| \) over all functions \( g \in L_2(\mathbb{Z} \times \mathbb{R}) \), satisfying (7.29).

**Proof:** Applying \( \mathcal{U}^* \) to \( g = \mathcal{W}x \) we have

\[
x(t)w(t-a) = \int_{\mathbb{R}} g(a, \omega)B^*(t - a, \omega)d\omega.
\]

Multiplying left and right by \( w(t-a) \) we obtain

\[
x(t)w^2(t-a) = \int_{\mathbb{R}} g(a, \omega)w(t-a)B^*(t - a, \omega)d\omega.
\]

Summation over \( a \) leads to

\[
x(t)\|w\|^2 = \sum_{a \in \mathbb{Z}} \int_{\mathbb{R}} g(a, \omega)w(t-a)B^*(t - a, \omega)d\omega,
\] (7.30)

or,

\[
x(t) = \frac{1}{\| w \|^2} \sum_{a \in \mathbb{Z}} \int_{\mathbb{R}} g(t - a, \omega)w(a)B^*(a, \omega)d\omega.
\]

Let \( \tilde{g} \) be a function satisfying (7.29). Taking the summation over \( t \) after multiplication by \( x^*(t) \) in (7.30), and assuming that interchanging summation and integral is allowed, leads to:

\[
\| x \|^2 \|w\|^2 = \sum_{t \in \mathbb{Z}} \sum_{a \in \mathbb{Z}} \int_{\mathbb{R}} \tilde{g}(a, \omega)x^*(t)w(t-a)B^*(t - a, \omega)d\omega
\]

\[
= \sum_{a \in \mathbb{Z}} \int_{\mathbb{R}} \tilde{g}(a, \omega)(\mathcal{W}x)^*(a, \omega)d\omega,
\]
From Theorem 7.4.2 we have, \( \|x\|^2 \|w\|^2 = \langle \mathcal{W}x, \mathcal{W}x \rangle \), thus:

\[
\sum_{a \in \mathbb{Z}} \int_{\mathbb{R}} \left( \mathcal{W}x - \tilde{g} \right)(a, \omega) \left( \mathcal{W}x \right)^*(a, \omega) d\omega = 0
\]

Thus \( \mathcal{W}x \) and \( \mathcal{W}x - \tilde{g} \) are orthogonal, so

\[
\|\tilde{g}\|^2 = \|\mathcal{W}x\|^2 + \|\mathcal{W}x\|^2 \geq \|\mathcal{W}x\|^2,
\]

with the equality if and only if \( \tilde{g} = \mathcal{W}x \)

From this theorem, a general reconstruction formula can be written as:

\[
\tilde{g} = \mathcal{W}x + v,
\]

where \( v \in \{ \mathcal{W}x \}^\perp \). Besides the reconstruction formula given in Theorem 7.4.3, we note that there are various ways to reconstruct the signal \( x \) from its CSWUS, as described in [76]. Consequently, the CSWUS is a redundant description of the signal \( x \). In fact, it is not necessary to know the complete CSWUS to reconstruct \( x \), as we will see next. The CSWUS can be written in convolutional form as:

\[
(\mathcal{W}x)(a, \omega) = (x \ast \tilde{k}(\cdot, \omega))(a).
\] (7.31)

So the CSWUS \( (\mathcal{W}x)(\cdot, \omega) \) can be seen as an analysis filter bank with analysis filter impulse responses \( \tilde{k}(\cdot, \omega), \omega \in \mathcal{J} \).

**Lemma 7.4.4** Let \( x \in \ell_p(\mathbb{Z}) \) and \( h \in \ell_1(\mathbb{Z}) \), \( p \geq 1 \). Then \( x \ast h \in \ell_p(\mathbb{Z}) \). Furthermore, \( \|x \ast h\|_p \leq \|h\|_1 \|x\|_p \).

**Proof:** Let \( x \in \ell_p(\mathbb{Z}) \) and \( h \in \ell_1(\mathbb{Z}) \), then

\[
\left( \sum_{n \in \mathbb{Z}} \left| (h \ast x)(n) \right|^p \right)^{1/p} = \left( \sum_{n \in \mathbb{Z}} \left( \sum_{m \in \mathbb{Z}} |h(m)||x(n - m)| \right) \right)^{1/p}
\]

\[
= \left( \sum_{n \in \mathbb{Z}} \left( \sum_{m \in \mathbb{Z}} |h(m)|^{1/p}|x(n - m)||h(m)|^{1-1/p} \right)^p \right)^{1/p}
\]

\[
\leq \left( \sum_{n \in \mathbb{Z}} \left( \sum_{m \in \mathbb{Z}} |h(m)||x(n - m)|^p \right) \left( \sum_{m \in \mathbb{Z}} |h(m)| \right)^{p-1} \right)^{1/p}
\]

\[
\leq \|h\|_1 \|x\|_p,
\]

thus \( x \ast h \in \ell_p(\mathbb{Z}) \) and \( \|x \ast h\|_p \leq \|h\|_1 \|x\|_p \).

So the convolution \( \ast \) defines a Banach algebra in \( \ell_p(\mathbb{Z}) \). The following Lemma states that a reconstruction is possible if \( \Psi(\cdot, \omega_0) \) is convolutionally invertible.
7.4 Localizing unitary transformations and filter banks

Lemma 7.4.5 (Wiener) [101] Let \( h \in \ell_1(\mathbb{Z}) \), with \( z \)-transform \( H(z) \). If \( \text{H}(z) \neq 0 \) for all \( z \in \mathbb{U} \), then there exists \( k \in \ell_1(\mathbb{Z}) \) such that \( h \ast k = \epsilon_0 \), that is,

\[
\forall_{z \in \mathbb{U}} \quad \frac{1}{H(z)} = \sum_{t \in \mathbb{Z}} k(t) z^{-t}.
\]

Applying the \( z \)-transformation to (7.31) we obtain

\[
Y(z, \omega) = X(z) \hat{K}(z, \omega), \quad z \in \mathbb{U}
\]

where \( y(a, \omega) = (\mathcal{W}x)(a, \omega) \), \( Y(z, \omega) \) and \( K(z, \omega) \) are the \( z \)-transforms of \( y(a, \omega) \) and \( \hat{k}(a, \omega) \) with respect to \( a \), respectively. Suppose now that there exists \( \omega_0 \in \mathbb{J} \) such that \( \hat{K}(z, \omega_0) \neq 0 \) for all \( z \in \mathbb{U} \), then it follows from Wiener’s Lemma that there exists \( h(\cdot, \omega_0) \in \ell_1(\mathbb{Z}) \) such that

\[
H(z, \omega_0) = \frac{1}{\hat{K}(z, \omega_0)} , \quad z \in \mathbb{U}
\]

So a reconstruction of \( x \) is obtained

\[
X(z) = Y(z, \omega_0) H(z, \omega_0), \quad z \in \mathbb{U}
\]

or, equivalently,

\[
x(t) = ((\mathcal{W}x)(\cdot, \omega_0) \ast h(\cdot, \omega_0))(t), \quad t \in \mathbb{Z}.
\]

We see that the condition \( \hat{K}(z, \omega_0) \neq 0 \) for a certain \( \omega_0 \in \mathbb{J} \), for all \( z \in \mathbb{U} \), is sufficient for the reconstruction of a signal \( x \) from its CSWUS. We shall prove that this condition is necessary for reconstruction of \( x \) from \( (\mathcal{W}x)(\cdot, \omega_0) \), for a fixed \( \omega_0 \). In fact, from Definition 7.4.1 we have

\[
(\mathcal{W}x)(a, \omega_0) = \langle x, \sigma^a k(\cdot, \omega_0) \rangle
\]

for all \( t \in \mathbb{Z} \), where

\[
k(t, \omega) = w(t) B^*(t, \omega_0).
\]

To reconstruct the signal \( x \) from this partial knowledge of its CSWUS we need to show that \( \{\sigma^l k(\cdot, \omega_0) | l \in \mathbb{Z} \} \) is a frame in \( \ell_2(\mathbb{Z}) \).

Theorem 7.4.6 The system \( \{\sigma^l k(\cdot, \omega_0) | l \in \mathbb{Z} \} \) is a frame in \( \ell_2(\mathbb{Z}) \) if and only if

\[
\hat{K}(z, \omega_0) \neq 0
\]

for all \( z \in \mathbb{U} \). Furthermore, if \( \{\sigma^l k(\cdot, \omega_0) | l \in \mathbb{Z} \} \) is a frame in \( \ell_2(\mathbb{Z}) \), then it is a Riesz basis in \( \ell_2(\mathbb{Z}) \).
**Proof:** See [73].

It is obvious that there is a similarity between an analysis filter bank, as shown in Fig. 7.4, and a subsampled CSWUS on $a = lL$, $l \in \mathbb{Z}$, $\omega = \omega_0, \ldots, \omega_{M-1}$, in $\mathbb{F}$, $M$ is a positive integer. Here $L$ is the ‘uniform’ sampling factor (a positive integer).

Such a discrete subset for subsampling is shown in Fig. 7.11.

**Remark 7.4.7** The choice of the sample values $\omega_0, \ldots, \omega_{M-1}$ is not free, it has to be such that a good choice of discreteness of the “pseudo-”inverse of $\mathcal{W}$ can be obtained to ensure a well-conditioned reconstruction problem.

In the case $L = M$, we say the local spectrum is critically sampled, in the case $L < M$ we say the local spectrum is oversampled. Suppose now that the spectrum of a signal $x$ is known for the distinct sample values of $\omega$, that is, $\omega_0, \ldots, \omega_{M-1}$, in $\mathbb{F}$ and for $a = lL$ ($l \in \mathbb{Z}$). For this partial knowledge of the spectrum CSWUS we have

\[
(\mathcal{W}x)(lL, \omega_m) = \langle x, \sigma^{lL} k(\cdot, \omega_m) \rangle,
\]

\[
k(t, \omega_m) = w(t)B^*(t, \omega_m), \quad m = 0, \ldots, M-1.
\]

The subsampled spectrum can be reformulated in a simple filtering scheme:

\[
(\mathcal{W}x)(lL, \omega_m) = \langle x, \sigma^{lL} k(\cdot, \omega_m) \rangle
\]

\[
= \sum_{t \in \mathbb{Z}} x(t) k^*(-lL + t, \omega_m)
\]

\[
= (k^*(\cdot, \omega_m) * x)(lL)
\]

\[
= (\downarrow L(a_m * x))(l),
\]
for \( l \in \mathbb{Z} \) and \( m = 0, \ldots, M - 1 \), with
\[
a_m(t) = \tilde{k}^*(t, \omega_m), \quad t \in \mathbb{Z},
\]
the impulse response of the analysis filter \( A_m \). The subsampled information of \( \mathcal{W}x \) defines an analysis filter bank; see Fig. 7.4, with the analysis filters \( A_m, m = 0, \ldots, M - 1 \).

We can reconstruct the signal \( x \) perfectly from its sampled spectrum \( \mathcal{W}x \) if and only if the system \( \{\sigma^L k(\cdot, \omega_m)\} \) is a frame in \( \ell_2(\mathbb{Z}) \), see previous section, that is, if there exist \( A > 0, B > 0 \) such that
\[
A \|x\|^2 \leq \sum_{l,m} |(\mathcal{W}x)(lL, \omega_m)|^2 \leq B \|x\|^2,
\]
for all \( x \in \ell_2(\mathbb{Z}) \). If this is the case, the frame operator \( A \) is identified by the subsampled \( \mathcal{W}x \):
\[
(Ax)(l, m) = (\mathcal{W}x)(lL, \omega_m), \quad l \in \mathbb{Z}, m = 0, \ldots, M - 1.
\]
The analysis operator \( \mathcal{J} = A^* A \) reads
\[
\mathcal{J}x = \sum_{m=0}^{M-1} \sum_{l \in \mathbb{Z}} (\mathcal{W}x)(lL, \omega_m) \sigma^L \tilde{a}_m^*, \quad x \in \ell_2(\mathbb{Z}).
\]
The filters \( \tilde{k}^*(\cdot, \omega_m), \quad m = 0, \ldots, M - 1 \), are causal and stable in \( \ell_2 \)-sense. The reconstruction of the signal \( x \) is given by
\[
x = \sum_{m=0}^{M-1} \sum_{l \in \mathbb{Z}} (\mathcal{W}x)(lL, \omega_m) \sigma^L \mathcal{J}^{-1} \tilde{a}_m^*,
\]
where we applied Lemma 7.3.8. Of course, the reconstruction of \( x \) from its subsampled \( \mathcal{W}x \) using filter banks can be realized with a synthesis filter bank as is shown in Fig. 7.5. Here the impulse response \( s_m, m = 0, \ldots, M - 1 \), satisfies
\[
s_m = \mathcal{J}^{-1} \tilde{a}_m^*.
\]

7.5 Linear constant coefficient difference equations

In this section we give reconstruction schemes of synthesis filter bank, given a (critically or oversampled) subsampled CSWUS. We start from a frame identified by the analysis filters \( A_m, m = 0, \ldots, M - 1 \) and its dual frame identified by the synthesis filters \( S_m, m = 0, \ldots, M - 1 \). The filters \( S_m, m = 0, \ldots, M - 1 \) are in general
non-causal. Furthermore, the transfer functions $A_m$ and $S_m$, $m = 0, \ldots, M-1$, are not necessarily rational filters. For realizability, this is obviously required.

Therefore, we start now with a set of transfer functions $A_m(z)$, $m = 0, \ldots, M-1$, not necessarily generating a frame. In order to derive necessary and sufficient conditions to obtain FIR or IIR filters, we introduce some variables which we will need, $u(l)$, $y(l)$ and $A(l)$, respectively,

$$u(l)(i) = x(lL + i) \quad (7.32)$$
$$y(l)(m) = (\mathcal{W}x)(lL, \omega_m) = \langle x, \sigma^{lL} a_m \rangle$$
$$A(l)(m, i) = \hat{h}^i(lL + i, \omega_m) = a_m(lL + i) \quad (7.33)$$

for $i = 0, \ldots, L-1$, $m = 0, \ldots, M-1$, $l \in \mathbb{Z}$. So for $x \in \ell_2(\mathbb{Z})$ we have $u \in \ell_2(\mathbb{Z}, \mathbb{C}^K)$, $y \in \ell_2(\mathbb{Z}, \mathbb{C}^M)$. Note that $H(z)$ is the polyphase transfer matrix obtained from the analysis filter bank with filters $\Psi_m$, $m = 0, \ldots, M-1$. From the previous definitions we have now

$$y(t) = \sum_{l \in \mathbb{Z}} A(t - l)u(l), \quad t \in \mathbb{Z}, \quad (7.34)$$

note that $A$ is causal in the sense that $A(l) = 0$ for $l < 0$. The analysis filter bank is BIBO stable since $w \in \ell_1(\mathbb{Z})$ and that the kernel $B$ is bounded, this means that $A \in \ell_1(\mathbb{Z}, \mathbb{C}^{M \times L})$. Now we give a definition of finite-order filtering in the multidimensional case.

**Definition 7.5.1** An input-output relation, linear constant matrix coefficient difference equation, of the form (7.34) is said to be of finite-order with order $N$ if there exist $P_0, \ldots, P_N$ in $\mathbb{C}^{M \times L}$ and $Q_1, \ldots, Q_N$ in $\mathbb{C}^{M \times M}$ such that

$$y(t) = \sum_{l=0}^{N} P_l u(t - l) - \sum_{l=1}^{N} Q_l y(t - l), \quad t \in \mathbb{Z}. \quad \square$$

The following theorem tells us when $A$, (7.33), leads to a finite-order filtering.

**Theorem 7.5.2** An input-output relation of the form (7.34) leads to a finite-order filtering with order $N$ if and only if there exist matrices $(E_n)_{n=1, \ldots, N}$ in $\mathbb{C}^{M \times M}$ such that

$$A(l) = \sum_{n=1}^{N} E_n A(l - n), \quad \text{for all } l > N.$$
Then in the notation of the definition 7.5.1

\[ P_l = A(l) - \sum_{n=1}^{l} E_n A(l - n), \quad \text{for } l = 0, 1, \ldots, N, \]

\[ Q_l = -E_l, \quad \text{for } l = 1, \ldots, N. \]

\[ \square \]

**Proof:** Suppose that there exist \((P_l)_{l=0,\ldots,N} \in \mathbb{C}^{L \times M}\) and \((Q_l)_{l=1,\ldots,N} \in \mathbb{C}^{L \times L}\) such that

\[ y(k) = \sum_{l=0}^{N} P_l u(k - l) - \sum_{l=1}^{N} Q_l y(k - l), \quad \text{for all } k \in \mathbb{Z}. \]

Substitution of (7.34) leads to

\[ \sum_{j \leq k} A(k - j) u(j) + \sum_{l=1}^{N} Q_l \left( \sum_{j \leq k-l} A(k - l - j) u(j) \right) = \sum_{l=0}^{N} P_l u(k - l), \]

for all \(k \in \mathbb{Z}\). Let \(k > N\) then we have

\[ \sum_{l=0}^{N} P_l u(k - l) = \sum_{j \leq k} A(k - j) u(j) + \sum_{l=1}^{N} \sum_{j \leq k} Q_l A(k - l - j) u(j) \]

\[ - \sum_{l=1}^{N} \sum_{k-l < j \leq k} Q_l A(k - l - j) u(j), \]

for all \(k > N\), from the causality of \(A\), and the substitution of \(j = k - l\) in the left hand term, we obtain

\[ \sum_{j \leq k} \left( A(k - j) + \sum_{l=1}^{N} Q_l A(k - l - j) \right) u(j) = \sum_{j=k-N}^{k} P_{k-j} u(j), \]

for all \(k > N\) and all \(u\) as defined in (7.32). So we have

\[ A(k - j) + \sum_{l=1}^{N} Q_l A(k - l - j) = P_{k-j}, \quad \text{for } k - N \leq j \leq k \]

and

\[ A(k - j) + \sum_{l=1}^{N} Q_l A(k - l - j) = 0, \quad \text{for } j < k - N. \]
Thus

\[ A(k) = - \sum_{l=1}^{N} Q_l A(k - l), \text{ for } k > N. \]

Conversely, suppose that there exist \((E_l)_{l=1, \ldots, N}\) such that

\[ A(k) = \sum_{l=1}^{N} E_l A(k - l), \text{ for } k > N. \]

Then we have

\[
y(k) = \sum_{l \leq k} A(k - l) u(l) \\
= \sum_{l \geq 0} A(l) u(k - l) \\
= \sum_{l=0}^{N} A(l) u(k - l) + \sum_{l > N} A(l) u(k - l) \\
= \sum_{l=0}^{N} A(l) u(k - l) + \sum_{j=1}^{N} \left( \sum_{l \leq k - N - 1} E_j A((k - j) - l) u(l) \right) \\
= \sum_{l=0}^{N} A(l) u(k - l) + \sum_{j=1}^{N} E_j \left( \sum_{l \leq k - N - 1} A((k - j) - l) u(l) \right) \\
= \sum_{l=0}^{N} A(l) u(k - l) + \sum_{j=1}^{N} E_j y(k - j) - \sum_{k - N \leq l \leq k - j} A((k - j) - l) u(l) \\
= \sum_{l=0}^{N} A(l) u(k - l) + \sum_{j=1}^{N} E_j y(k - j) - \sum_{j=1}^{N} E_j \left( \sum_{j \leq l \leq N} A(l - j) u(k - l) \right) \\
= \sum_{l=0}^{N} (A(l) - \sum_{j=1}^{k} E_j A(l - j)) u(k - l) + \sum_{j=1}^{N} E_j y(k - j) \\
= \sum_{l=0}^{N} Q_l u(k - l) - \sum_{j=1}^{N} Q_j y(k - j),
\]

which leads to an input-output relation as defined in Definition 7.5.1.

**Example 7.5.3** Consider the Fourier system

\[
B(t, \omega) := \frac{1}{\sqrt{2\pi}} e^{-j\omega t},
\]
\( \omega \in [0, 2\pi), t \leq 0, \) and the window
\[
w(t) := \begin{cases}
\sqrt{2\pi r^{-t}} & \text{if } t \leq 0 \\
0 & \text{otherwise.}
\end{cases}
\]
where \( 0 < r < 1. \) Then \( A(k) \) is given by
\[
A(k) = \begin{pmatrix}
\tau_{kL} e^{-jkL\omega_0} & \cdots & \tau_{kL} e^{-jkL-1e-j(kL+L-1)\omega_0} \\
\tau_{kL} e^{-jkL\omega_1} & \cdots & \tau_{kL} e^{-jkL-1e-j(kL+L-1)\omega_1} \\
\vdots & \cdots & \vdots \\
\tau_{kL} e^{-jkL\omega_{M-1}} & \cdots & \tau_{kL} e^{-jkL-1e-j(kL+L-1)\omega_{M-1}}
\end{pmatrix}
\]
for \( k \in \mathbb{Z} \). There is the following finite-order property
\[A(k) = CA(k-1), \quad k > 0\]
or,
\[A(k) = C^kA(0) = C^kP_0,\]
where
\[
C = \text{diag}(\tau^L e^{-jkL\omega_0}, \ldots, \tau^L e^{-jkL\omega_{M-1}})
\]
and
\[
P_0 = \begin{pmatrix}
1 & \tau e^{-j\omega_0} & \cdots & \tau^{L-1} e^{-j(L-1)\omega_0} \\
1 & \tau e^{-j\omega_1} & \cdots & \tau^{L-1} e^{-j(L-1)\omega_1} \\
\vdots & \vdots & \cdots & \vdots \\
1 & \tau e^{-j\omega_{M-1}} & \cdots & \tau^{L-1} e^{-j(L-1)\omega_{M-1}}
\end{pmatrix}
\]
\[
\square
\]
For the analysis filter bank, finite-order filtering means that the filters \( A_m(z), m = 0, \ldots, M - 1, \) are elements of \( \mathbb{C}(z) \). The transfer functions \( A_m(z), m = 0, \ldots, M - 1, \) are FIR or IIR filters if and only if the condition in Theorem 7.5.2 is satisfied. The filters are stable because \( a_m \in \ell_1(\mathbb{Z}), m = 0, \ldots, M - 1. \) For the existence of perfect reconstruction the analysis filter bank has to be injective. Next, we derive reconstruction formulas for these analysis filter banks. We want to reconstruct a signal from its subsampled CSWUS straightforwardly from the analysis filter bank.

The filtering scheme in terms of the required analysis filters which we shall use is shown in Fig. 7.12. The transfer \( S(z) \) is the synthesis filter transfer matrix. Now
we can formulate the problem to find $S(z)$. Let the input-output relation (7.34) have finite-order $N$. So there exist $(E_n)_{n=1,\ldots,N}$ in $\mathbb{C}^{M \times M}$ such that

$$A(l) = \sum_{n=1}^{N} E_n A(l-n), \ l > N,$$

and

$$y(t) = \sum_{l=0}^{N} P_l u(t-l) - \sum_{l=1}^{N} Q_l y(t-l), \ t \in \mathbb{Z}, \quad (7.35)$$

Applying the $z$-transformation to (7.35) we obtain

$$A(z) = \left( I + \sum_{l=1}^{N} z^{-l} Q_l \right)^{-1} \left( \sum_{l=0}^{N} z^{-l} P_l \right), \ z \in \mathbb{U}. $$

So for the stability of $A(z)$ it follows that all the zeros of the determinant of the polynomial matrix

$$B(z) := I + \sum_{l=1}^{N} z^{-l} Q_l, $$

have to be within the unit circle $\mathbb{U}$. We recall that a filter bank, Fig. 7.12, has the property of perfect reconstruction if there exists a rational $L \times M$-matrix $K(z)$ with

\[\text{Figure 7.12: } M \text{ channel } L \text{ subsampled filter bank.}\]
entries in $\mathbb{C}(z)$ such that $K(z)T(z) = I$, for every $z \in \mathbb{U}$, where

$$T(z) := \sum_{l=0}^{N} z^{-l}P_{l}, \quad z \in \mathbb{U} \tag{7.36}$$

We note that

$$T(z) = B(z)A(z),$$

thus for $K(z)$, with $K(z)T(z) = I$, we have

$$K(z)B(z)A(z) = I.$$

So $K(z)B(z)$ defines the reconstruction scheme. Necessary and sufficient conditions to obtain causal perfect reconstruction are given in the next two theorems.

**Theorem 7.5.4** Suppose $L = M$ and let $T(z)$, see (7.36), be of full rank on the unit circle. There exists a unique causal synthesis filter for a perfect reconstruction if and only if $P_0$ is invertible.

**Proof:** Let $T(z)$ be of full rank on $\mathbb{U}$. Thus $(T(z))^{-1}$ exists for each $z \in \mathbb{U}$. Suppose there exists a causal perfect reconstruction synthesis filter $K(z)$, such that

$$K(z)B(z)A(z) = K(z)T(z) = I.$$

The general form of the causal $K(z)$ is given by:

$$K(z) = \sum_{n=0}^{\infty} z^{-n}K(n), \quad z \in \mathbb{U}$$

From $K(z)T(z) = I$, we have:

$$\left(\sum_{n=0}^{\infty} z^{-n}K(n)\right) \left(\sum_{l=0}^{N} z^{-l}P_{l}\right) = I,$$

for all $z \in \mathbb{U}$. Consequently, we have:

$$K(n)P_0 = \begin{cases} I & \text{if } n = 0 \\ -\sum_{l=0}^{n-1} K(l)P_{n-l} & \text{if } n \geq 1. \end{cases}$$

Thus $P_0$ is invertible; with $P_0^{-1} = K(0)$, and

$$K(n) = -\sum_{l=0}^{n-1} K(l)P_{n-l}K(0).$$
Conversely, suppose $P_0$ is invertible, then we have from (7.35):

$$u(t) = P_0^{-1}y(t) + \sum_{l=1}^{N} P_0^{-1}Q_ly(t - l) - \sum_{l=1}^{N} P_0^{-1}P_ly(t - l),$$

for all $t \in \mathbb{Z}$.

\[\square\]

**Theorem 7.5.5** Suppose $L < M$ and let $T(z)$, see (7.36), be of full rank on the unit circle. There exists a causal synthesis filter for a perfect reconstruction if and only if $P_0$ is injective.

**Proof:** The proof for the oversampled case is similar to the critical one. Here we can use any left inverse of $P_0$ for $K(0)$.

As a consequence of a causal synthesis filter bank satisfying perfect reconstruction, there exist an integer $N_1 > 0$ and matrices $C_l, l = 0, \ldots, N_1$, in $\mathbb{C}^{L \times M}$ and $D_l$, $l = 1, \ldots, N_1$, in $\mathbb{C}^{L \times L}$ such that

$$u(t) = \sum_{l=0}^{N_1} C_l y(t - l) - \sum_{l=1}^{N_1} D_l u(t - l), \quad t \in \mathbb{Z},
\tag{7.37}$$

for all $u$ and $y$ related by (7.35). This reconstruction formula gives a synthesis matrix $S(z)$.

$$S(z) = K(z)B(z) = \left( I + \sum_{l=1}^{N_1} z^{-l}D_l \right)^{-1} \left( \sum_{l=0}^{N_1} z^{-l}C_l \right),$$

for $z \in \mathbb{U}$. Stability of the synthesis filter bank is required, meaning that all the poles of $K(z)B(z)$ have to be within the unit circle $\mathbb{U}$.

**Problem 7.5.6** Find a filter $K(z)$, $L \times M$ transfer rational matrix, such that

$$K(z)\Upsilon(z) = I$$

and $K(z)B(z)$ is stable.

If such a $K(z)$ is found then we obtain a stable perfect reconstruction filter bank, Fig. 7.12.

**7.5.1 Perfect reconstruction in the critical case**

In the critical case, that is, $L = M$, we obtain for $K(z)$ the inverse of $\Upsilon(z)$, which is unique, and the stability of $K(z)B(z)$ requires the singularities of $\Upsilon(z)$ (points
7.5 Linear constant coefficient difference equations

where $\Upsilon(z)$ is not injective) to be within the unit circle. The reconstruction formula, in this case, exists if $A(0) = P_0$ is invertible, namely

$$u(t) = P_0^{-1}y(t) + \sum_{l=1}^{N_1} P_0^{-1}Q_l y(t - l) - \sum_{l=1}^{N_1} P_0^{-1}P_l u(t - l),$$

for $t \in \mathbb{Z}$, or, using the $z$-transformation

$$\left( I + \sum_{l=1}^{N} z^{-l}P_0^{-1}P_l \right) U(z) = P_0^{-1} \left( I + \sum_{l=1}^{N} z^{-l}Q_l \right) Y(z),$$

for $z \in \mathbb{U}$. For the stability we need that all the zeros of the determinant of $I + \sum_{l=1}^{N} z^{-l}P_0^{-1}P_l, z \in \mathbb{C}^*$, are within the unit circle. Then we obtain for the synthesis transfer $S(z)$

$$S(z) = \left( I + \sum_{l=1}^{N} z^{-l}P_0^{-1}P_l \right)^{-1} P_0^{-1} \left( I + \sum_{l=1}^{N} z^{-l}Q_l \right).$$

This leads to a stable perfect reconstruction.

**Theorem 7.5.7** [76] Suppose $L = M$ and let $\Upsilon(z)$ be defined as in (7.36). If all the zeros of $\det(\Upsilon(z))$ are within the unit circle and $P_0$ is invertible, then there exists a unique stable causal perfect reconstruction for the filter bank, Fig. 7.12. So it is required that all the zeros of $\det(\Upsilon(z))$ and $\det(B(z))$ are within the unit circle in order to obtain a stable perfect reconstruction filter bank.

7.5.2 Perfect reconstruction in the oversampled case

In the oversampled case, that is, $L < M$, a left inverse $K(z)$ of $\Upsilon(z)$ is not unique. In Theorem 7.5.7 the matrix $K(z)$ is unique, which is not the case here. So a theorem as Theorem 7.5.7 can not be expected. For $K(z)$ we can take the pseudo-inverse which is the unique solution of the pointwise least squares estimate (Projection Theorem) [61, 76]. In general, in order to be able to solve Problem 7.5.6

$$K(z) \Upsilon(z) = I,$$

where $\Upsilon(z)$ is given by (7.36). It is necessary that $\Upsilon(z)$ is of full rank on the unit circle. If this is satisfied then every left inverse of $\Upsilon(z)$ can be used as a solution of
the above problem. If \( K(z) \) is a left inverse of \( \Upsilon(z) \) then the transfer of the synthesis filter becomes

\[
S(z) = K(z)B(z).
\]

If \( \Upsilon(z) \) is of full rank on the unit circle \([76]\), then

\[
\| (\Upsilon(z)^*\Upsilon(z))^{-1} \Upsilon(z)^* \| = \min \{ \| K(z) \| : K(z)\Upsilon(z) = I, K(z) \in \mathbb{C}^{L \times M} \},
\]

for all \( z \in \mathbb{U} \).

**Theorem 7.5.8** [76] Let \( \Upsilon(z) \) be defined as in (7.36). If all the singularities of \( \Upsilon(z) \) lie within the unit circle \( \mathbb{U} \), (zeros of \( \det(\Upsilon(z)^*\Upsilon(z)) \)), then the synthesis filter

\[
S(z) = (\Upsilon(z)^*\Upsilon(z))^{-1} \Upsilon(z)^*B(z)
\]

is the minimal, minimum norm point wise, stable perfect reconstruction for the filter bank, Fig. 7.12.

However, the condition that the singularities are within the unit circle is only a sufficient condition to guarantee perfect reconstruction, but not a necessary one. In fact, there are many ways to obtain left inverses of \( \Upsilon(z) \), which may lead to a stable perfect reconstruction, for instance, the one based on the so-called *Smith Form* \([54, 76]\).

If \( P_0 \) is injective, we find as a recursive reconstruction formula

\[
u(t) = K_0 \left( y(t) + \sum_{l=1}^{N_1} Q_l y(t-l) - \sum_{l=1}^{N_2} P_l u(t-l) \right),
\]

for \( t \in \mathbb{Z} \), where \( K_0 \) is a left inverse of \( P_0 \). We can choose for \( K_0 \) the pseudo-inverse \((P_0^*P_0)^{-1}P_0^* \) of \( P_0 \). Applying the \( z \)-transformation we obtain

\[
\left( I + \sum_{l=1}^{N} z^{-l} K_0 P_l \right) U(z) = K_0 \left( I + \sum_{l=1}^{N} z^{-l} Q_l \right) U(z).
\]

For a stable perfect reconstruction filter transfer for the synthesis filter bank, see Fig. 7.12, it is required that all the zeros of \( \det \left( I + \sum_{l=1}^{N} z^{-l} K_0 P_l \right) \) are within the unit circle. In this case, the synthesis filter becomes

\[
S(z) = \left( I + \sum_{l=1}^{N} z^{-l} K_0 P_l \right)^{-1} K_0 \left( I + \sum_{l=1}^{N} z^{-l} Q_l \right).
\]

The synthesis filters based on the injectivity of \( P_0 \) are causal, but the synthesis filters based on the minimum norm of left inverses of \( \Upsilon(z) \) (pseudo-inverse) may yield non-causal filters. We note that these synthesis filters operate at the lower rate.
7.6 Kautz and Laguere filter banks

In Part I, we have seen how to optimize the parameters of orthogonal expansions, including Laguerre and Kautz expansions. In this section we present how the Laguerre and Kautz systems can be used in perfect reconstructing filter banks. We define the kernel for these systems, and the corresponding CSWUS for special choices of the window function \( w \). Further, we derive the linear constant matrix coefficient difference equation and show that the reconstruction satisfies a simple FIR filtering scheme.

From Definition 5.2.1, we recall the definition of a Kautz system: let \((\lambda_n)_{n \in \mathbb{N}_0}\) be a sequence within the unit disc; with \(\lambda_n\)'s not necessarily mutually distinct. Define the natural number \( j_n \) by \( j_n = \sum_{l=0}^{n-1} \delta_{\lambda_n \lambda_l} \) (\(\delta_{\lambda_n \lambda_l}\) is equal to 1 if \(\lambda_n = \lambda_l\), otherwise zero). Let \( f_n \in \ell_2(\mathbb{N}_0) \) be defined as follows \( f_n(k) = \binom{k}{j_n} \lambda_{n-j_n} \), \( k, n \in \mathbb{N}_0 \), recall that by definition that \( \binom{k}{l} = 0 \) if \( l > k \). The system \( \{f_n|n \in \mathbb{N}_0\} \) is linearly independent. Further, let \( \{g_n|n \in \mathbb{N}_0\} \) be the result of the Gram-Schmidt orthonormalization procedure applied to \( \{f_n|n \in \mathbb{N}_0\} \). The obtained orthonormal system \( \{g_n|n \in \mathbb{N}_0\} \) is called a Kautz system.

**Theorem 7.6.1** [86] A Kautz system is total in \( \ell_2(\mathbb{N}_0) \) if and only if

\[
\sum_{n=0}^{\infty} (1 - |\lambda_n|) = \infty.
\]

\( \square \)

A Laguerre system is the particular case where all the \(\lambda_n\)'s are equal, so \( j_n = n \). For a Kautz system as described above with different \(\lambda_n\)'s, where all \(\lambda_n\)'s are mutually distinct, we consider \( J = \mathbb{N}_0 \), \( K = \mathbb{Z} \), and for the kernel-function

\[
B(t, l) = \begin{cases} 
g_l(-t) & \text{if } t \leq 0, \\
0 & \text{otherwise.} \end{cases}
\]

for \( l \in \mathbb{N}_0 \). Then we have

\[
\langle B(\cdot, l), B(\cdot, l') \rangle = \delta_{ll'}
\]

\[
\langle B(t, \cdot), B(t', \cdot) \rangle = \delta_{tt'}.
\]

So the system \( \{B(-t, \cdot)|t \in \mathbb{N}_0\} \) is orthonormal in \( \ell_2(\mathbb{N}_0) \).
Definition 7.6.2 The causal Sliding-Window Kautz Transformation (SWKT) is defined by

\[
(G_w x)(a, l) = \sum_{t \in \mathbb{Z}} x(t + a)w(t)B(t, l), \ a \in \mathbb{Z}, l \in \mathbb{N}_0,
\]

with \(w \in \ell_1(\mathbb{Z})\) and causal.

For a window \(w\) we consider

\[
w(t) = r^{-t},
\]

for \(t \leq 0\), where \(0 < r < 1\). We note that this choice of the window \(w\) guarantees the stability of the analysis filters. We consider the subsampling obtained by \(l = 0, \ldots, M - 1\), \(a = kL, k \in \mathbb{Z}, M \geq L > 0\). The analysis filters are given by:

\[
a_m(t) = w(-t)B(-t, m) = w(-t)g_m(t), \ m = 0, \ldots, M - 1.
\]

For the transfer function in the \(z\)-domain we refer to (5.6) in Chapter 5. So for a finite-order filtering, see (7.34), we obtain for \(A(l)\)

\[
A(l) = r^L P(l) D, \ l \in \mathbb{N}_0,
\]

where

\[
P(l)(i, j) = g_i(lL + j), \ i = 0, \ldots, M - 1, \ j = 0, \ldots, L - 1
\]

\[
D = \text{diag}(1, r, \ldots, r^{L-1}).
\]

The system \(\{g_l | l \in \mathbb{N}_0\}\) is the result of the Gram-Schmidt orthonormalization procedure applied to the system \(\{f_l | l \in \mathbb{N}_0\}\) with \(f_l(t) = \chi_l^t\). Thus \(P(l)\) can be written as \(P(l) = GQ(l)\), where \(G\) is the lower triangular matrix induced by the Gram-Schmidt orthonormalization procedure and

\[
Q(l)(i, j) = \chi_{i+j}^L,
\]

for \(i = 0, \ldots, M - 1; \ j = 0, \ldots, L - 1\). Further, we can write \(Q(l) = E^l Q(0)\), where \(E = \text{diag}(\chi_0^L, \ldots, \chi_{M-1}^L)\). Thus \(A(l) = CA(l - 1)\), for \(l \in \mathbb{N}\), where \(C = r^L GEG^{-1}\) and \(A(0) = GQ(0)D\). This leads to the first-order input-output filtering

\[
y(t) = A(0)u(t) + Cy(t - 1), \ t \in \mathbb{N}_0.
\]

This analysis filtering relation is stable because the zeros of \(\text{det}(I - z^{-1}C)\) are within the unit circle. This input-output relation has a perfect reconstruction if and only if
A(0) is injective. The matrix \( A(0) \) is injective if and only if \( Q(0) \) is injective. \( Q(0) \) is injective if and only if there are \( M \lambda_n \)'s, \( 0 \leq n \leq M - 1 \), mutually distinct, which is given. Thus the input-output relation has always perfect reconstruction

\[
U(z) = KA \left( I - z^{-1}C \right) Y(z),
\]

where \( KA(0) = I \). This reconstruction is of first order FIR which is always stable. As an application, we refer to [19] where the Kautz transformation is used to interpret the critical bands in the auditory system.

Suppose now that the \( \lambda_n \)'s are equal, which is the Laguerre system case. Then we can put \( A(l) = s^{1L}P(l)D, \ l \in \mathbb{N}_0 \). In this case, we can show that \( P(l) = FP(l - 1) \), where \( F \) is a lower triangular matrix [16, 76]. Also in this case we have a stable FIR first-order perfect reconstruction with the same results as in the Kautz case.

The hybrid cases (a mix of Kautz and Laguerre) can be treated in a similar way. First, we write \( A \) as an upper block matrix with block matrices on the diagonal of ranks \( j_n \) with entries as in Laguerre case for \( M = j_n \). Then we can use the procedure as outlined above for the Laguerre case for each block matrix.

### 7.7 Discussion

Signal descriptions based on Fourier Transformation are generalized by replacing the kernel by kernels of unitary integral transformations. In combination with a suitable choice of window functions we obtained the (causal) windowed unitary transformation having the same properties as the Sliding-Window Discrete-Time Fourier Transformation [76]. For some sorts of subsampling we implemented the subsampling in analysis filter banks. Synthesis is directly obtained using frame theory. However, not every frame can be associated with a filter bank which can be used in practice. Indeed, the Fourier images of the filters have to be rational functions.

Another very interesting point is stability. In fact stability, in the frame concept means that a frame yields a well-conditioned reconstruction. This reflects that the analysis operator is bounded and bounded from below. In contrast, the stability in the filter bank concept means that the black box transfer systems associated with the filter bank are BIBO-stable.

Stable perfect reconstruction filter banks are constructed under some conditions on the window and the kernel functions. Further, the Kautz and the Laguerre transformations are presented as particular cases of the CSWUT. These transformations have practical interpretations: in the case of Kautz transformation with poles close to the unit circle we have a discrete-time discrete-frequency description, where the poles have the interpretation of frequency-samples. The description in this case is close to the case of Fourier transformation. In the case of Laguerre transformation, we have a polynomial decomposition by means of the CSWUT.
In contrast to FIR, IIR filters usually do not profit from the noble identities: application does in general lead to simpler (in terms of number of operations) filtering schemes. Therefore, using an IIR filter bank, a low-order filtering scheme is desirable to limit the number of operations. The Kautz and Laguerre transformations have this property: the analysis filter bank can be realized as a cascade of first-order IIR filters; the synthesis bank is a low-order FIR filter. Furthermore, the method described in this chapter for reconstruction guarantees the design of causal systems.
Chapter 8

Non-uniformly downsampled filter banks

The problem of perfect reconstructing non-uniformly downsampled filter banks is considered in this chapter. This problem can be reformulated to a uniformly downsampled filter bank thus allowing the usual analysis. Of special interest are those filter banks where the output of the analysis bank has a direct interpretation, for example, the sliding-window Fourier transformation or the wavelet transformation. The concept of the sliding-window Fourier transform can be extended by replacing the Fourier transformation by an arbitrary unitary transformation. As an example the sliding-window Kautz transformation is considered, which extends in a certain sense the uniform case considered in the previous chapter.

8.1 Non-uniform versus uniform

Analysis/synthesis filter banks are used to decompose signals into separate subbands. Classically, these decompositions have uniform structures, or uniform-based tree structures, that is, tree structures based on two-band decompositions. However, sometimes these decompositions do not allow a good match to requirements from practice. For example, for the subband coding of speech/audio signals, the ideal decomposition should match the critical bands of the ear [82, 103]. These bands are neither uniform nor representable in a uniform-based tree structure. The desired struc-

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{localizing_grids.png}
\caption{Localizing grids of a uniform and a non-uniform downsampling scheme, where \(a\) represents the original (time) domain and \(\omega\) the transfer domain (often frequency).}
\end{figure}
Figure 8.2: A non-uniform filter bank related to a Kautz system.

ture for decomposition is purely non-uniform. Some design methods for non-uniform perfect reconstruction filter banks have been proposed in [2, 30, 51, 59, 60, 70]. However, there is still a need for an efficient and simple design method. Here we present an approach which is maybe the starting point for implementing such a method.

Let us start from the two grids presented in Fig. 8.1. The grid shown in the left-hand side of Fig. 8.1 corresponds to a four channel uniform filter bank; with downsampling factor four. This filter bank satisfies the property that the sum of the stream ratio per channel \((1/4 + 1/4 + 1/4 + 1/4)\) is greater or equal to one, which is a necessary condition for the injectivity of the analysis filter bank. The right-hand side of Fig. 8.1 shows a grid which corresponds to a non-uniform filter bank, for example the one shown in Fig. 8.2; with downsampling factors 8, 8, 4 and 2. This filter bank satisfies also the property that the sum of the stream ratio per channel \((1/8 + 1/8 + 1/4 + 1/2)\) is greater or equal to one, which is of course a necessary condition for the injectivity of the non-uniform downsampled analysis filter bank. The difference here is that the subband signals are non-uniformly determined. In other words, the compression differs per channel. This gives the possibility to subsample a localized spectrum dependent on its characteristics. Applying manipulations to the downsamplers in the non-uniform downsampled filter bank, that is, adding downsamplers and upsamplers, the analysis filter bank becomes uniform. For example, the analysis part of the filter bank shown in Fig. 8.2 becomes the 8-channel filter bank in Fig. 8.5. In Section 8.2, we show how this can be done and in general how a non-uniform filter bank can be linked to a uniform one.

Consider the analysis bank shown in Fig. 8.3 consisting of the \(M\) filters with impulse responses \(a_m (m = 0, \ldots, M-1)\) cascaded with the downsamplers \(L_m\). The input signal is \(x\) and the output signals (subband signals) are \(y_m (m = 0, \ldots, M-1)\) and these output signals each have their specific sample frequency \(f_s/L_m\) with \(f_s\) the sample frequency of the input signal \(x\). The filters \(a_m\) are assumed to be stable, that is, \(a_m \in \ell_1(\mathbb{Z})\), and causal. The filters are of FIR or IIR type: the \(z\)-transform of \(a_m\)
is a rational function of $z$ denoted by $A_m$. For the output signals we can write

$$y_m(l) = \sum_{t \in \mathbb{Z}} x(t) a_m(lL_m - t) = \langle x, \sigma^{lL_m} a_m^* \rangle$$

(8.1)

From frame theory, see previous chapter, we know that the reconstruction of the signal $x$ is possible if and only if the set $\{\sigma^{lL_m} a_m^* | m = 0, \ldots, M-1, l \in \mathbb{Z}\}$ constitutes a frame in $\ell_2(\mathbb{Z})$, that is,

$$\exists_{A > 0, B > 0} \forall x \in \ell_2(\mathbb{Z}) \quad A ||x||^2 \leq \sum_{l,m} ||\langle x, \sigma^{lL_m} a_m^* \rangle||^2 \leq B ||x||^2.$$  

(8.2)

It is in general not easy to check whether a given filter bank constitutes a frame using the previous condition. A necessary condition to obtain a frame from this analysis bank is that $\rho \geq 1$, where $\rho$ is the oversampling factor defined by

$$\rho := \sum_{m=0}^{M-1} \frac{1}{L_m}.$$  

This is intuitively clear since, otherwise, the data stream per unit of time at the output of the analysis bank is less than the data stream per unit of time at its input. The condition $\rho \geq 1$ will be proved more rigorously later on, when we link the non-uniform analysis filter bank to a uniform one. In general, as we have seen in the previous chapter, the reconstruction by a frame is not unique. We have a unique reconstruction if and only if the set $\{\sigma^{lL_m} a_m^* | m = 0, \ldots, M-1, l \in \mathbb{Z}\}$ constitutes a Riesz basis in $\ell_2(\mathbb{Z})$ in which case we have $\rho = 1$.

If the system $\{\sigma^{lL_m} a_m^* | m = 0, \ldots, M-1, l \in \mathbb{Z}\}$ is a frame in $\ell_2(\mathbb{Z})$, we can reconstruct the signal $x$ using the analysis operator $\mathcal{F}$

$$\mathcal{F} x = \sum_{l,m} \langle x, \sigma^{lL_m} a_m^* \rangle \sigma^{lL_m} a_m^* = \sum_{l,m} y_m(l) \sigma^{lL_m} a_m^*$$

Figure 8.3: Non-uniformly downsampled analysis bank.
which is invertible under the frame assumption. So the reconstruction becomes

\[ x = \sum_{l,m} y_m(l) \mathcal{J}^{-1} \sigma^{l,m} \tilde{a}_m^*. \]

It would lead to an efficient computation of the synthesis filters if \( \mathcal{J} \) and \( \sigma^{l,m} \) commute for \( m = 0, \ldots, M - 1 \). However, in general they do not. From Lemma 7.3.8, we obtain that \( \mathcal{J} \) and \( \sigma^{kL} \) commute for \( L = \text{lcm}(L_0, \ldots, L_{M-1}) \), \( k \in \mathbb{Z} \). Also, the computation of \( \mathcal{J}^{-1} \) is not simple; see Section 7.3 in the previous chapter. To solve this problem, one mostly uses a Neumann series to approximate \( \mathcal{J}^{-1} \):

\[ \mathcal{J}^{-1} = \frac{2}{B + A} \sum_{k=0}^{\infty} (I - \frac{2}{B + A} \mathcal{J})^k, \]

where \( A, B \) are the frame bounds. A good approximation of \( \mathcal{J}^{-1} \) is then obtained if \( A \approx B \). This reconstruction obtained by an approximation of \( \mathcal{J}^{-1} \) would be not a perfect one.

What we really would like is that we can reconstruct the signal \( x \) (or a delayed version thereof) using causal and stable filters. This is not guaranteed even if the system \( \{ \sigma^{l,m} \tilde{a}_m^* \mid m = 0, \ldots, M - 1, l \in \mathbb{Z} \} \) is a frame: if the reconstruction exists this means that we can do the reconstruction using pattern functions in \( \ell_2(\mathbb{Z}) \) but this is insufficient to meet our goals, namely to achieve a simple design method for the reconstruction using causal filters (for example, Laguerre and Kautz filters).

### 8.2 Equivalent uniformly downsampled filter bank

We can redraw the analysis bank of Fig. 8.3 in the way shown in Fig. 8.4 where \( L \) is the least common multiple \( L = \text{lcm}(L_0, \ldots, L_{M-1}) \), cf. [56]. If we draw a time-channel grid indicating the sampling instants at the output of the analysis bank, \( L \) is the number on the time-axis after which the sampling pattern repeats itself. Note that the part left of the dashed line in Fig. 8.4 consists of causal filters. We consider now the reconstruction of this uniformly downsampled analysis bank.

First we introduce some additional variables. We have the outputs of the uniform analysis filter bank represented by the part before the dashed line in Fig. 8.4

\[ y_{m,j}(l) = \sum_{t \in \mathbb{Z}} x(t) a_m(tL - jL_m - t), \]

for \( m = 0, \ldots, M - 1, j = 0, \ldots, L/L_m - 1 \). Next we have the input and output vector signals \( u(l) \in \mathbb{C}^L \) and \( y(l) \in \mathbb{C}^{M_L} \) defined by

\[ u(l) = (x(lL), x((lL - 1)), \ldots, x((lL - L + 1)))^\top \]  
\[ y(l) = (y_{0,0}(lL), y_{0,1}(lL), \ldots, y_{M-1,L/L_{M-1}-1}(lL)))^\top \]  

(8.3)  
(8.4)
Equivalent uniformly downsampled filter bank

Figure 8.4: Equivalent uniformly downsampled filter bank; here we used the following notation: $D_m = L/L_{m-1}$, $m = 1, \ldots, M$, $A_{m,1} = z^{-a_{m,1}}A_{m-1}$, $a_{m,l} = (l - 1)L_{m-1}$, $l = 1, \ldots, D_m$, $m = 1, \ldots, M$.

for $l \in \mathbb{Z}$, and the analysis matrices

$$\{A(l)\} (n, i) = a_m(lL - jL_m + i),$$

where $l \in \mathbb{Z}$, $i = 0, \ldots, L - 1$, $n(m, j) = j + \sum_{k=0}^{m-1} L/L_k$, $m = 0, \ldots, M - 1$ and $j = 0, \ldots, L/L_m - 1$. Note that $A(l)$ is causal: $A(l) = 0$ for $l < 0$. The matrix sequence $A(l)$ is no more than a stack containing the impulse responses of the filters of Fig. 8.4. The matrix $A$ has dimensions $\rho L \times L$. From these definitions we can write the processing in the analysis bank as a convolution

$$y(l) = \sum_{k \leq l} A(l - k)u(k). \quad (8.5)$$

Note that $\sum_{k \leq l}$ expresses the causality of the analysis bank. Using these manipulations for the non-uniformly downsampling analysis part of the filter bank drawn in Fig. 8.2, we obtain the filter bank in Fig. 8.5. Here we have $L = 8 = \text{lcm}(8, 8, 4, 2)$.

For a finite-order filtering we have the relation; see Definition 7.5.1:

$$y(n) = \sum_{k=0}^{N} P_k u(n - k) - \sum_{k=1}^{N} Q_k y(n - k). \quad (8.6)$$
In the $z$-domain this can be written as

$$
\left( I + \sum_{k=1}^{N} z^{-k} Q_k \right) Y(z) = \left( \sum_{k=0}^{N} z^{-k} P_k \right) U(z),
$$

(8.7)

The analysis bank is stable if and only if the zeros of $\det(I + \sum_{k=1}^{N} z^{-k} Q_k)$ are within the unit circle. This last holds if and only if the filters $a_m$ are stable which was assumed from the outset.

For a non-uniformly upsampled synthesis filter bank, we can apply the same manipulations, described before, to link it to a uniformly upsampled synthesis filter bank. In this case, we obtain the filter bank shown in Fig. 8.6. The right part of the dashed line in Fig. 8.6 is an $pL$-channel uniform synthesis bank. We note that the left part of the dashed line in Fig. 8.6 can be considered reciprocal to the right part of the dashed line in Fig. 8.4. Explicitly, let $A_{nu} = F \circ A_u$ denote$^1$ the analysis bank of the non-uniformly downsampling filter bank, where $A_u$ represents the uniformly downsampling analysis filter bank in the left-hand side of Fig. 8.4 and $F$ represents the system in the right-hand side of Fig. 8.4. Similarly, for the non-uniformly upsampled synthesis filter bank we define $S_{nu} = S_u \circ G$, where $S_u$ represents the uniform synthesis filter bank in the right-hand side of Fig. 8.6 and $G$ represents the system in

$^1$ $\circ$ denotes the composition, so $B \circ A$ means the composition of the systems $A$ and $B$, and the system $A$ is applied first.
8.2 Equivalent uniformly downsampled filter bank

the left-hand side of Fig. 8.6. Thus,

\[ S_{nu} \circ A_{nu} = S_u \circ G \circ F \circ A_u = S_u \circ A_u. \]

From this equality, we see that the reconstruction problem can be approached from two different ways. The first approach is to hold the non-uniform scheme, see Fig. 6.4, which is obvious. However, this straightforward approach loses the nice properties, to write the input-output relation in matrices.

\[ X_{out}(z) = \sum_{m=0}^{M-1} \sum_{l=0}^{L_m-1} S_m(z) A_m(\Omega_{L_m}^l z) X_m(\Omega_{L_m}^l z). \]

Of course, we have perfect reconstruction if the coefficient of \( X_m(z) \) is a delay and if the rest-term equals zero for all possible input signals. For design perspectives, it is desired to have simple reconstruction schemes. Therefore, we suggest the second approach, namely, to consider the reconstruction scheme by the equivalent form. This approach manipulates the non-uniform problem to a uniform one. Next, we give a reconstruction scheme by the second approach. In Section 8.4, we define the concept of best non-perfect reconstruction, where the perfect reconstruction condition is slightly weakened.

Reconstruction

Consider the input-output relation (8.7) in the \( z \)-domain. It is clear that there is a perfect reconstruction if \( P_0 \) is injective. For a stable perfect reconstruction it is required that the zeros of \( \text{det}(I + \sum_{k=1}^{N} z^{-k} K_0 P_k) \) are within the unit circle, where \( K_0 P_0 = I \). If both requirements are met we have a causal stable perfect reconstruction by

\[ u(n) = K_0 \left[ y(n) + \sum_{k=1}^{N} Q_k y(n - k) \right] - \sum_{k=1}^{N} K_0 P_k u(n - k) \quad (8.8) \]

or, in the \( z \)-domain, the synthesis filter is given by

\[ S(z) = \left( I + \sum_{k=1}^{N} z^{-k} K_0 P_k \right)^{-1} K_0 \left( I + \sum_{k=1}^{N} z^{-k} Q_k \right). \]

In general, including the non-causal case, \( K_0 \) is non-unique. Only in the critically sampled case the reconstruction is unique. In the oversampled case we can use, for example, the pseudo-inverse of \( P_0 \). Given the analysis bank and the downsampling scheme (thus \( P_m \) and \( Q_m \)) and requiring causal reconstruction, we only have freedom to choose \( K_0 \) under the constraint \( K_0 P_0 = I \).
That the reconstruction is non-unique in the oversampled case has several advantages. We can exploit this non-uniqueness for instance to obtain simple reconstruction filters in terms of the number of operations required to do the reconstruction, for minimizing (or shaping) the noise in the reconstructed signal caused by the noise usually introduced on the signal $y(k)$ or to reduce the coefficient sensitivity \cite{14, 31, 32}. We can write the reconstructed signal $r$ by the causal reconstruction formula

$$ r(n) = \sum_{k=0}^{\infty} S_k y(n - k) $$

with the definition of $r(n)$ similar to (8.3) and where $S_k$ are the reconstruction matrices. The noise power $\sigma_r^2$ in the reconstructed signal $r$ caused by additive white noise with variance $\sigma_y^2$ introduced at the output of the analysis bank (that is, on $y_{l,j}$) can be simply written in terms of these matrices as

$$ \sigma_r^2 = \frac{\sigma_y^2}{L} \sum_{k=0}^{\infty} \text{trace} \{S_k S_k^*\}. $$

This form is especially suitable if the reconstruction is of the FIR-type.
8.3 Examples

Frame bounds of a non-uniform frame

To test if the non-uniform system \{\sigma_{l,m}^* a_m^* | n = 0, \ldots, M - 1, l \in \mathbb{Z}\} constitutes a frame in \ell_2(\mathbb{Z}) we can use the alias component matrix \(H(z)\) of the uniform analysis bank

\[
\{H(z)\}_{n,i} = \frac{1}{\sqrt{L}}(\Omega_L^i z)^j L_m A_m(\Omega_L^j z)
\]

where \(n = j + \sum_{k=0}^{m-1} L/L_m, \Omega_L = e^{2\pi i / L}, i = 0, \ldots, L - 1, j = 0, \ldots, L/L_m - 1.\)

The system \(\{\sigma_{l,m}^* a_m^* | n = 0, \ldots, M - 1, l \in \mathbb{Z}\}\) is a frame in \(\ell_2(\mathbb{Z})\) if and only if there exist \(A > 0, B > 0\) such that

\[
AI \leq H(z)^* H(z) \leq BI,
\]

for all \(z\) on the unit circle with \(0 \leq \arg(z) \leq 2\pi/L.\) From this we can immediately infer that a necessary condition is that \(\rho L \geq L\) and thus \(\rho \geq 1.\)

8.3 Examples

A special and attractive case of filter banks is obtained in case the behaviour of the analysis bank can be interpreted easily, for instance, if the output \(y\) is a subsampled sliding-window Fourier transform (SWFT) of the input \(x.\) In that case, the index \(m (m = 0, \ldots, M - 1)\) associated with the filters immediately takes on a physical interpretation in terms of a center frequency (SWFT). Thus the time-channel grid becomes a time-frequency grid (SWFT). As we have seen in the previous chapter, the idea of a sliding-window Fourier transformation can be extended to a sliding-window unitary transformation. As a specific example we considered the Kautz transformation as defined in Definition 7.6.2, Section 7.6 of the previous chapter.

The simplest case of a Sliding-Window Kautz Transformation (SWKT) occurs if the window function is a causal exponential sequence. Furthermore, we take the first \(M\) basis functions of the Kautz system with arbitrary poles \(\lambda_m.\)

**Theorem 8.3.1** For the SWKT based on a causal exponential window, the first \(M\) basis functions of the Kautz system with arbitrary poles \(\lambda_m,\) and an arbitrary non-undersampled and non-uniform downsampling, there exists a causal stable synthesis bank which is of first-order FIR (in the downsampled domain).

In terms of the finite-order filtering and the previous notation we have as in the uniform case \(\mathcal{N} = 1\) and \(A(k) = CA(k - 1)\) for \(k \geq 1.\)

We thus have that the SWKT has a local energy interpretation, and that the poles \(\lambda_m\) and the subsampling grid can be chosen freely. Furthermore, the noise analysis
is quite simple as a consequence of the first-order reconstruction, namely

$$\sigma_r^2 = \frac{\sigma_0^2}{L} \text{trace} \{ K_0(I + CC^*)K_0^* \}. $$

The causal reconstruction for which $\sigma_r^2$ is minimal is called the minimal causal dual frame and is attained, see [61], for

$$K_0 = (P_0^*(I + CC^*)^{-1}P_0)^{-1}P_0^*(I + CC^*)^{-1}. $$

Consider the Kautz transformation with poles as shown in Fig. 8.7. The ordering of these poles is taken in order of increasing radius. Consequently, the first filter in the windowed Kautz analysis bank has the broadest bandwidth and complex-conjugated poles occur sequentially. The parameter of the exponential window sequence is indicated by the circle.

![Figure 8.7: The poles of the Kautz transformation (crosses) and window parameter (circle).](image)

In Fig. 8.8 we have plotted the amplitude transfers of the filters performing the windowed Kautz transformation as defined by the poles and window parameter of Fig. 8.7.

![Figure 8.8: Amplitude transfers of the analysis bank.](image)
In principle, we can select a quite arbitrary downsampling scheme in cascade with these analysis filters. For perfect reconstruction we have critical sampling or oversampling.

As an example we take the downsampling scheme as shown in Fig. 8.9 in the form of a time-frequency grid. Horizontally we have the time-axis and vertically the center frequencies of our analysis filters. The channel index number is shown on the right. The crosses indicate the samples taken after the analysis. The downsamplers per channel are taken (nearly) inverse proportional to the bandwidth with the restriction of a repeating sampling pattern after 24 samples ($L = 24$). There are 25 samples within the fundamental cell of length 24, which means there is a slight oversampling with a factor $p = 25/24$:

$$\frac{1}{4} + \frac{1}{6} + \frac{1}{6} + \frac{1}{8} + \frac{1}{8} + \frac{1}{12} + \frac{1}{12} + \frac{1}{24} = \frac{25}{24}$$

As stated earlier, the reconstruction is of first-order FIR in the downsampled domain. This means that we have to determine two matrices ($S_1$ and $S_2$) of sizes $24 \times 25$ to do the reconstruction. These matrices can be determined numerically.

Other time-frequency grids are possible as well. We could use for instance a uniform downsampling scheme with decimators with factor 8 after each analysis filter. We still have perfect reconstruction by first-order FIR, but now the two reconstruction matrices are of size $8 \times 8$. Uniform downsampling gives the minimum possible delay. However, such scheme is not attractive from the point of view of additive noise. Noise (for example signal quantization noise) introduced at the output of the analysis bank results in more noise power (nearly a factor 5 in this example):

$$\frac{\sigma_{r,nu}^2}{\sigma_{r,mu}^2} = 4.8083,$$

in the reconstruction for the uniform downsampling scheme than for the downsampling scheme of Fig. 8.9.
8.4 Best non-perfect reconstruction

The input-output relation of a non-uniform filter bank is given by:

\[ X_{\text{out}}(z) = \sum_{m=0}^{M-1} \frac{1}{L_m} \sum_{l=0}^{L_m-1} S_m(z)A_m(\Omega_{l,m}^I z)X_{\text{in}}(\Omega_{l,m}^I z). \]

From the equivalent form we can rewrite the input-output relation as,

\[ X_{\text{out}}(z) = \frac{1}{L} \sum_{m=0}^{M-1} \sum_{l=0}^{L-1} \gamma(l, m)S_m(z)A_m(\Omega_{L}^I z)X_{\text{in}}(\Omega_{L}^I z) \]

\[ = \sum_{l=0}^{L-1} T_l(z)X_{\text{in}}(\Omega_{L}^I z), \]

where

\[ \gamma(l, m) = \sum_{d=0}^{L_m-1} \Omega_{L}^{-ldL_m} = \begin{cases} \frac{L}{L_m} & \text{if } l = 0 \mod \frac{L}{L_m} \\ 0 & \text{otherwise.} \end{cases} \]

This is a simplified representation, since it factorizes \( X_{\text{in}} \) using a single \( \Omega_{L} \) instead of \( \Omega_{l,m} \), \( m = 0, \ldots, M - 1 \). It is clear that we have perfect reconstruction, if the overall distortion transfer function \( T_0(z) \) satisfies

\[ T_0(z) = \sum_{m=0}^{M-1} \frac{1}{L_m} S_m(z)A_m(z) = cz^{-n_0}, \]

for some constant \( c \) and an integer \( n_0 \), and if the alias transfer functions \( T_l(z) \) satisfy

\[ T_l(z) = \frac{1}{L} \sum_{m=0}^{M-1} \gamma(l, m)S_m(z)A_m(\Omega_{L}^I z) = 0. \]

From a practical perspective, these perfect reconstruction conditions are too restrictive. In practice, it is sufficient to design the filter bank such that, the overall distortion transfer function \( T_0(z) \) satisfies

\[ \| T_0(z) - cz^{-n_0} \| < \Delta_1, \]
and the alias transfer functions $T_i(z)$ satisfy

$$||T_i(z)|| < \Delta_2,$$

where $\Delta_1$ and $\Delta_2$ are small numbers, representing the accuracy of the filter bank. The number $\Delta_2$ may represent, for example, the stopband attenuation. The norm here can be taken as $||G(z)|| = \sup_{z \in \Omega} |G(z)|$. In this problem formulation, the analysis filters are taken fixed, and the synthesis filters have to be found, (of course, for fixed synthesis filters to find analysis filters we can do the same). This problem is similar to the $\mathcal{H}_\infty$ optimization problem. For more details we refer to [69].

### 8.5 Discussion

We have reviewed concepts of non-uniformly downsampled filter banks from frame theory and filter bank theory. The elegance of the considered unitary transformation is its simple implementation of both the analysis and synthesis bank and its large degree of freedom of choosing a time-frequency grid.

For good numerical properties, that is, immunity to signal quantization noise and coefficient quantization, the proper choices of the poles in the Kautz transformation and the time-frequency grid are undoubtedly interconnected. This issue has to be explored further.
Part III

Analysis by domain segmentations
Chapter 9

Introduction to Part III

In this introduction, we clarify the issues we want to treat in this part. We consider the segmentation problems for discrete-time signals, and we present mathematical solutions. Here we do not deal with perceptive models, since then other parameters are needed to be taken into account. Let us start with a real example. We consider a sampled version $x$ of the speech signal corresponding to the sentence: ”You’re cool!” uttered by a child, sampled at a rate of $11025 \text{Hz}$. In Fig. 9.1 we present the amplitude and the Fourier spectrum of this signal. The Fourier spectrum, that is, the frequency spectrum, shows the global distribution of the energy of the signal $x$ as a function of frequency. Intuitively, the peaks in the figure show the frequencies most strongly present in the energy of the signal.

![Figure 9.1: Recording of ”You’re cool!” uttered by a child.](image)

For coding purposes it is sometimes interesting to segment the signal into small parts to characterize the local properties of the speech. Applying this idea to the speech signal above, the sentence ”You’re cool!” can be split intuitively into three parts according to breaks of silence, that is, approximately corresponding to the words you-re-cool!, see Fig. 9.2. So the discrete-time domain $I$ of the signal $x$ is segmented into three parts $I_1, I_2$ and $I_3$. In each of these intervals the signal $x$ cor-
responds with a part of the sentence, with distinctive properties. In this example we have for $I, I_1, I_2$ and $I_3$ the discrete intervals $[1, 7830]$, $[1, 2500]$, $[2501, 3400]$ and $[3401, 7830]$. The parts of the segment are given by $\chi_{I_1} x$, $\chi_{I_2} x$ and $\chi_{I_3} x$. Time-frequency representation of the signal $x$ gives also a segmentation of the signal in the time × frequency domain. This is motivated by the uncertainty principle that states that the energy spread of a function and its Fourier transform cannot be simultaneously arbitrarily small. Gabor [43] has used time-frequency atoms ‘waveforms’ as decomposition of signals to measure the time-frequency information content. Gabor argued that such decompositions are closely related to our sensitivity audio. In Fig. 9.3 we show a spectrogram of the utterance ”You’re cool!” giving the energy distribution over the time-frequency axis. The energy is presented by the grey-level, dark for the high values of instantaneous energy and light for the lower values.

From Fig. 9.3 it is obvious to choose the intuitive time-segmentation. However, it may be efficient to decompose the signal in five partitions instead of the suggested three, since there are time-intervals where the signal has a very small contribution of the signal energy, for example $[1, 500]$ and $[2000, 2800]$. Further, we remark that some patterns are showing harmonic dependence on the frequency.

Having the domain-segmentation $I = I_1 \cup I_2 \cup I_3$ of the signal $x$, it is desired to approximate or compress the segment $\chi_{I_i} x$, $i = 1, 2, 3$, in a few number of elements of a basis in $\ell_2 (I_i)$. The local bases desired to use in audio, are required to constitute interpretable characteristics related to this special type of signals, for instance the discrete Fourier series. Given a discrete interval $J = [a, b]$, we define such a basis.
The index $l$ of $\phi_l$ corresponds to the phase $\frac{2l + 1}{b - a + 1} \pi$.

As we have seen from the plots of the frequency spectrum and the local spectrum, see Fig 9.1 and Fig 9.3, not all frequencies are significantly present in the signal $x$. To describe $x$ efficiently in a compressed representation in terms of the basis $\{\phi_l[J] | l = 0, \cdots, b - a + 1\}$ by

$$\phi_l(t; J) = \chi_J(t) \frac{\sqrt{2}}{\sqrt{b - a + 1}} \cos \left( \frac{2l + 1}{b - a + 1} \pi \left( t + \frac{1}{2} - a \right) \right), \quad t \in J.$$ 

The norm is defined in $\ell_2(I)$.

For the signal "You're cool!" we used 985 ($N = 985, N_1 = 491, N_2 = 50, N_3 = 444$) coefficients depicted in Fig. 9.4 that led to an approximation containing 90% of the energy of $x$, $\|x - x_N\|^2 / \|x\|^2 \approx 0.1$. For

\footnote{The norm is defined in $\ell_2(I)$.}
Figure 9.4: The chosen coefficients in the approximation of $x$ using local cosine basis.

In this approximation we propose the following method: For the discrete-time interval $I_1$, the frequency domain has been split according to the corresponding parts $F_k$ in the local spectrum, (the so-called Heisenberg boxes, see Fig. 9.5), $F_k$ is a discrete interval. For instance, for $I_1$ we divided the Fourier spectrum into 12 intervals, $F_k$, $k = 1, \ldots, 12$. In each of these intervals we have chosen the number of the largest coefficients leading to the 90% accuracy, in total 491 coefficients for the interval $I_1$. In the second interval $I_2$, we have chosen 50 largest coefficients. We see in the local spectrum, Fig. 9.3, that only frequencies between 500Hz and 2000Hz are significantly present in $I_2$. In the third interval $I_3$ we did the same as in the first time-interval $I_1$, resulting in 444 coefficients. In Fig. 9.6 we present the approximation containing 90% of the energy of $x$, its Fourier spectrum and the difference between the original $x$ and its approximation, that is, the error in approximation. In Fig. 9.7 we see that the local spectrum of the approximation is very similar to the local spectrum of the original signal, except for the transition domains, where we observe a clear ‘discontinuity’ in the transition from $I_1$ to $I_2$ and from $I_2$ to $I_3$, which is perceptible. The approximation apparently has all the frequencies present, but it shows the so-called boundary artifacts. In our segmentation $I = I_1 \cup I_2 \cup I_3$ we have chosen breaks of silence as boundaries of the intervals $I_1, I_2$ and $I_3$. For the starting points of $I_1$ and $I_2$ we chose the mid-times of each interval, and for the starting point of $I_3$ we chose the end time of $I_2$.

\[ \text{Werner Heisenberg lived from 1901 to 1976: attended school in Munich and he entered the University of Munich. There he studied physics under Sommerfeld. After completing his undergraduate course he continued study for his doctorate presenting his doctoral dissertation in 1923 on turbulence in fluid streams. Heisenberg did important work in nuclear and particle physics, but his most important work was in quantum mechanics.} \]
one of these intervals the basis functions (in the case of cosine basis) are all different from zero. So to get an accurate approximation also at the boundaries, we need to use many coefficients so that the artifact is suppressed.

If we consider the equidistant segmentation for $I = I_1 \cup I_2 \cup I_3$, ($I_1 = [1, 2610]$, $I_2 = [2611, 5220]$ and $I_3 = [5221, 7830]$), where the boundaries are in this case not corresponding to the breaks of silence, then we need in general more coefficients to obtain roughly an approximation containing the same quantity (90%) as the approxi-
mation in the first case. In Fig. 9.8 we give the plots of the approximation containing 90% of the energy of $x$, its Fourier spectrum and the error between the original $x$ and its approximation, in the equidistant three-interval segmentation. In this approximation we need 1150 coefficients, $\left(N = 1150, N_1 = 530, N_2 = 200, N_3 = 420\right)$ to present 90% of the energy of $x$, $\|x - x_N\|^2 \approx 0.1$.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure9_7.png}
\caption{Time-frequency representation of the approximated reconstruction of the speech signal $x$.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure9_8.png}
\caption{Signal approximation, frequency spectrum and the difference between the original signal and its reconstruction in the case of equidistant segmentation.}
\end{figure}
The plot in Fig. 9.9 shows the local spectrum of the approximation of \( x \), using the three-interval approximation. We observe here also a clear discontinuity in the transition domains. Comparing the two approximations we see that the first approximation is doing better in the sense that it uses a smaller number of coefficients than the second approximation. However, both of them show clear discontinuities in the transition domains.

\[ \text{Figure 9.9: Time-frequency representation of the approximated reconstruction of the speech signal } x \text{ in the case of equidistant segmentation.} \]

In Fig. 9.1 we see that the signal \( x \) contains intervals of silence, (these intervals contain a negligible quantity of the signal energy). So if we take this information into account and divide \( I \) into five segments, namely \( I_1 = [1, 500], I_2 = [501, 2000], I_3 = [2001, 2800], I_4 = [2801, 3400] \) and \( I_5 = [3401, 7830] \). For \( I_1 \) and \( I_3 \) we approximate \( x \) by zero. For the discrete-time intervals \( I_2, I_4 \) and \( I_5 \), we approximate \( x \) as previously. In this case we find that 978 coefficients are enough for an approximation containing more than 92% of the energy of \( x \) (\( N = 978 \)), that is, \( \| x - x_N \|^2 / \| x \|^2 < 0.08 \). The coefficients needed per segment are \( N_1 = N_3 = 0, N_2 = 428, N_4 = 50 \) and \( N_5 = 500 \). In Fig. 9.10 we show the plots of the approximation containing 92% of the energy of \( x \), its Fourier spectrum and the error between the original \( x \) and its approximation, in the case of five segments. The plot in Fig. 9.11 shows the local spectrum of the approximation of \( x \), using the five-interval approximation. We still observe a clear discontinuity in the transition domains. Comparing all these approximations, we see that the last approximation, five interval segmentation, is the most accurate approximation. However, all show clear discontinuity in the transition domains.
9.1 Goals

Summarizing the items playing a role above, we come to the following problems:

1. Given a signal $x$ with a time-domain $I$, for instance audio or speech, how can $x$ be segmented in local domains according to an interpretable criterion (for
example extraction of formants or sines/harmonics)? And if this is possible, what is the segmentation \( \cup_k I_k \) of \( I \)?

2. How to suppress the artifacts at the transition domains?

3. What are suitable bases to decompose signals in their local domains?

4. What are the criteria to choose the approximation in the local domains?

### 9.2 Outline

In part III we intend to review the techniques used in local representations. Namely, in Chapter 10, we elaborate on the second and third problems, and give some suggestions for the other problems. In Chapter 11, we consider the artifact reduction in the multidimensional case which can be useful, for example, for video. We discuss the generalization from the one-dimensional case by taking the tensor product, and we derive a non-separable case by hexagonal plane decomposition.

“I learned optimism from Sommerfeld, mathematics at Göttingen, and physics from Bohr.”

*Heisenberg*
160 Introduction to Part III
Chapter 10

Orthogonal transformations by 1-D domain segmentations

In the one-dimensional time domain represented by the real line, instead of disjointly segmenting, we decompose the real line into partly overlapping intervals. Each of these intervals is associated with an orthogonal projection, in which the overlap allows for some smoothness at the interval boundaries. In this way, we arrive at a resolution of the identity established by the set of mutually orthogonal projections. Thus, the one-dimensional signal space can be decomposed into a direct sum of mutually orthogonal subspaces that are images of the signal space under the orthogonal projection associated with the interval decomposition of the real line. For each of these subspaces, we can find bases such that a local orthogonal decomposition of signals can be carried out. In this chapter, we describe these orthogonal projections to reduce the boundary artifacts. We give conditions for local orthogonal bases for the corresponding local subspaces, we introduce redundancy by using frames instead of bases, and we consider how to choose the interval segmentation.

In Section 10.1, we consider the discrete-time case. This is the most interesting case from a practical point of view, and can also be seen as a sampled version of the continuous-time case. The continuous-time case is only shortly considered in Section 10.2, because it is sufficiently discussed in literature, cf. [3] whereas the discrete-time case is not. In Section 10.3, we address another aspect in local analysis, namely, the optimal interval decomposition where we have to base the localization upon. We conclude the chapter by a discussion in Section 10.4.

10.1 Lapped orthogonal segmentation for discrete signals

In this section we deal with artifact reduction for discrete-time signal segmentation. The artifacts as observed in the introduction of this part, see Figures 9.7, 9.9 and 9.11, can be reduced by orthogonal projections which are smooth versions of the projections obtained by multiplication by the characteristic function $\chi$. Now the role of the characteristic function is replaced by a function which has a smooth transition at
Orthogonal transformations by 1-D domain segmentations

the boundaries [3]. The analysis is as follows. Divide the discrete-time domain into two parts, for instance the set \( \mathbb{Z} \) into \( \mathbb{Z}_0^+ \) and \( \mathbb{Z}^- \), and define an orthogonal projection \( P \) identified by \( \mathbb{Z}_0^+ \) in such a way that \( P \) maps a ‘smooth’ signal \( x \) (for example, \( \sum_{t \in \mathbb{Z}} |x(t) - x(t - 1)| \) is small) in a signal \( Px \) which has ‘smooth’ transition in the cutoff of the time domain, and such that:

\[
\text{support} (Px) \cap \text{support} ((I - P)x) \subseteq [-\epsilon, \epsilon], \quad \epsilon \in \mathbb{N}_0,
\]

where \( -\frac{1}{2} \) is taken the symmetry point from \( \mathbb{Z}_0^+ \) into \( \mathbb{Z}^- \) (cutoff point). Thus the orthogonal projections \( P \) and \( Q := I - P \) divide \( x \) into two parts with a smooth transition in the cutoff point \(-\frac{1}{2}\). Using time-shifting we can move the cutoff to every point in the time domain still resulting in two orthogonal projections preserving the same smoothness and overlapping properties. Suppose we divide \( \mathbb{Z} \) into two discrete intervals \((-\infty, a) \) and \([a, \infty) \), \( a \in \mathbb{Z} \), then \( P_a := T^{-a}PT^a \) and \( Q_a := I - P_a \) are two orthogonal projections with cutoff point \( a - \frac{1}{2} \) and

\[
\text{support} (P_a x) \cap \text{support} (Q_a x) \subseteq [a - \epsilon, a + \epsilon).
\]

Now we are able also to define orthogonal projections for finite-length intervals. Indeed, let \( I = [a, b] \) be a discrete-time interval, \( a, b \in \mathbb{Z}, \ a \leq b \). if the orthogonal projections \( P_a \) and \( Q_{b+1} \) commute, then \( P_{a,b} := P_a Q_{b+1} \) is an orthogonal projection, where the support of \( P_{a,b} x \) is given by

\[
\text{support} (P_{a,b} x) \subseteq [a - \epsilon, b + \epsilon').
\]

The overlap parameters \( \epsilon, \epsilon' \) are identified by the projections \( P_a \) and \( Q_{b+1} \), respectively. Systematically, this way, it is possible to decompose \( \ell_2(\mathbb{Z}) \) in a direct sum of \( \ell_2(\mathbb{Z}) \), where the cutoff points are \( \frac{1}{2} + a_k \).

Having described these ideas roughly, we now turn to a more explicit description of this method. Define on \( \ell_2(\mathbb{Z}) \) the two orthogonal projections \( P_{[x_0^+, \epsilon, s^+]}, P_{[z^-, \epsilon, s^-]} \):

\[
\begin{align*}
\left( P_{[x_0^+, \epsilon, s^+]} x \right)(t) &= w^2(t; \epsilon) x(t) + s^+ w(t; \epsilon) w(-1 - t; \epsilon) x(-1 - t), \\
\left( P_{[z^-, \epsilon, s^-]} x \right)(t) &= w^2(-1 - t; \epsilon) x(t) + s^- w(-1 - t; \epsilon) w(t; \epsilon) x(-1 - t),
\end{align*}
\]

where the window function \( w[\epsilon] \) is satisfying the so-called power-complementarity condition

\[
\begin{align*}
w^2(t; \epsilon) + w^2(-1 - t; \epsilon) &= 1, \quad \epsilon \in \mathbb{N}_0, \\
\text{support}(w[\epsilon]) &= [-\epsilon, \infty), \\
\text{and} \\
s^- = -s^+ &\in \{-1, 1\}
\end{align*}
\]
10.1 Lapped orthogonal segmentation for discrete signals

It is clear that \( w(t; \epsilon) = 1 \), for \( t > \epsilon \). We note that for \( \epsilon = 0 \), we have \( w[0] = \chi_{Z_0^+} \), and the orthogonal projections \( \mathcal{P}[Z_0^+, \epsilon, s^+] \), \( \mathcal{P}[Z_-^-, \epsilon, s^-] \) become multiplication by the corresponding characteristic functions. The window function \( w[\epsilon] \) guarantees the smooth transition in the cutoff \( -\frac{1}{2} \): the projections \( \mathcal{P}[Z_0^+, \epsilon, s^+] \) and \( \mathcal{P}[Z_-^-, \epsilon, s^-] \) divide the signal into two parts which have a smooth transition in \( -\frac{1}{2} \). An example of a window \( w[\epsilon] \) satisfying these conditions, is given by:

\[
w(t; \epsilon) = \begin{cases} 1 & \text{if } t \geq \epsilon, \\ 
\sin \left( \frac{\pi}{2} \cdot \frac{-2(\epsilon^2 - \epsilon) + (4\epsilon + 1)(\epsilon + 1)^2}{(2\epsilon + 1)^2} \right) & \text{if } -\epsilon \leq t < \epsilon, \\ 0 & \text{if } t < -\epsilon. 
\end{cases} \quad (10.1)
\]

For an interval \( I = [a, b] \), \( a + \epsilon_a \leq b - \epsilon_b \), the orthogonal projection \( \mathcal{P}_{[I, \epsilon_a, \epsilon_b, s_a, s_b]} \), \( s_a, s_b \in \{-1, 1\}, \epsilon_{ab} \in \mathbb{N}_0 \times \mathbb{N}_0 \), is defined by:

\[
\mathcal{P}_{[I, \epsilon_a, \epsilon_b, s_a, s_b]} = \mathcal{T}^{-a} \mathcal{P}_{[Z_0^+, \epsilon_a, s_a]} \mathcal{T}^{a-1} \mathcal{P}_{[Z_-^-, \epsilon_b, s_b]} \mathcal{T}^{b+1},
\]

where \( \mathcal{T}^{a-b+1} = \mathcal{T}^{-b-1} \mathcal{T}^a \). The definition of \( \mathcal{P}_{[I, \epsilon_a, \epsilon_b, s_a, s_b]} \) makes sense because the projections \( \mathcal{T}^{-a} \mathcal{P}_{[Z_0^+, \epsilon_a, s_a]} \mathcal{T}^a \) and \( \mathcal{T}^{-b-1} \mathcal{P}_{[Z_-^-, \epsilon_b, s_b]} \mathcal{T}^{b+1} \) are orthogonal and commute:

\[
\mathcal{T}^{-a} \mathcal{P}_{[Z_0^+, \epsilon_a, s_a]} \mathcal{T}^{a-1} \mathcal{P}_{[Z_-^-, \epsilon_b, s_b]} \mathcal{T}^{b+1} = \mathcal{T}^{-b-1} \mathcal{P}_{[Z_-^-, \epsilon_b, s_b]} \mathcal{T}^{b+a} \mathcal{P}_{[Z_0^+, \epsilon_a, s_a]} \mathcal{T}^a,
\]

where commutativity requires that \( a + \epsilon_a \leq b - \epsilon_b \). The explicit form of the orthogonal projection \( \mathcal{P}_{[I, \epsilon_a, \epsilon_b, s_a, s_b]} \), \( \epsilon_{ab} = (\epsilon_a, \epsilon_b) \in \mathbb{N}_0 \times \mathbb{N}_0 \), is given by:

\[
\left( \mathcal{P}_{[I, \epsilon_a, \epsilon_b, s_a, s_b]} x \right)(t) = w(t; I, \epsilon_{ab}) \left[ w(t; I, \epsilon_{ab}) x(t) + s_a w(2a - 1 - t; I, \epsilon_{ab}) x(2a - 1 - t) + s_b w(2b + 1 - t; I, \epsilon_{ab}) x(2b + 1 - t) \right]. \quad (10.2)
\]

Here \( w[I, \epsilon_{ab}] \) can be written in terms of \( w[\epsilon] \), as follows:

\[
w(t; I, \epsilon_{ab}) = w(t - a; \epsilon_a) w(b - t; \epsilon_b), \quad a - \epsilon_a \leq t \leq b + \epsilon_b, \quad (10.3)
\]

satisfying the power-complementarity condition:

\[
w^2(t; I, \epsilon_{ab}) = w^2(2a - 1 - t; I, \epsilon_{ab}) + w^2(2b + 1 - t; I, \epsilon_{ab}) = 1, \quad (10.4)
\]

for all \( t \in \text{support}(w[I, \epsilon_{ab}]) = [a - \epsilon_a, b + \epsilon_b] \).
Having the smooth orthogonal projections \( \mathcal{P}_{[I, s_{a}, s_{b}]} \) and \( \mathcal{P}_{[J, s_{c}, s_{d}]} \) for adjacent intervals \( I = [a, b] \) and \( J = [b + 1, c] \), for the subspaces \( \mathcal{P}_{[I, s_{a}, s_{b}]} \ell_{2}(Z) \) and \( \mathcal{P}_{[J, s_{c}, s_{d}]} \ell_{2}(Z) \) to be perpendicular, it is required that:

\[
\mathcal{P}_{[I, s_{a}, s_{b}]} \mathcal{P}_{[J, s_{c}, s_{d}]} = 0. \tag{10.5}
\]

This is the case if \( \epsilon_{b+1} = \epsilon_{b}, \ s_{b+1} = -s_{b} \). The proof uses also \( \epsilon_{a} \leq b \leq \epsilon_{b} \leq b + 1 + \epsilon_{b+1} \leq c - \epsilon_{c} \), and

\[
w(t; I, \epsilon_{ab}) = w(t - a; \epsilon_{a})w(b - t; \epsilon_{b}), \quad a - \epsilon_{a} \leq t \leq b + \epsilon_{b},
\]

\[
w(t; J, \epsilon_{b+1,c}) = w(t - b - 1; \epsilon_{b})w(c - t; \epsilon_{c}), \quad b + 1 - \epsilon_{b} \leq t \leq c + \epsilon_{c}.
\]

The property (10.5) gives room for further refinement in the time domain and can be used to find an optimal segmentation. Having the product of the projections \( \mathcal{P}_{[I, s_{a}, s_{b}]} \mathcal{P}_{[J, s_{c}, s_{d}]} \) is orthogonal, it can be proved that

\[
w^{2}(t; I, \epsilon_{ab}) + w^{2}(t; J, \epsilon_{b+1,c}) = w^{2}(t; I \cup J, \epsilon_{ac}), \tag{10.6}
\]

\[
\mathcal{P}_{[I, s_{a}, s_{b}]} + \mathcal{P}_{[J, s_{c}, s_{d}]} = \mathcal{P}_{[I \cup J, s_{a}, s_{b}, s_{c}]} \tag{10.7}
\]

In this way, given a partition \( \{ I_{k} \mid k \in Z \} \) of \( Z \), where the discrete-time interval \( I_{k} = [a_{k}, a_{k+1}] = \{ n \in Z \mid a_{k} \leq n < a_{k+1} \}, k, a_{k} \in Z \), and the projections \( \mathcal{P}_{[I_{k}, s_{a}, s_{b}]} \), \( \epsilon_{k} = (\epsilon_{a_{k}}, \epsilon_{a_{k+1}}) \in \mathbb{N} \times \mathbb{N}_{0}, s_{k} = (s_{a_{k}}, s_{a_{k+1}}, a_{k+1} - a_{k}), s_{a_{k}} = -s_{a_{k}} \), \( s_{a_{k}}, s_{a_{k+1}} \in \{-1, 1\} \), we decompose the signal space \( \ell_{2}(Z) \) into a direct sum of mutually orthogonal subspaces, that is, they form the resolution of the identity. In other words, the following decomposition:

\[
\ell_{2}(Z) = \bigoplus_{k \in Z} \mathcal{P}_{[I_{k}, s_{a}, s_{b}]} \ell_{2}(Z)
\]

that is

\[
\sum_{k \in Z} \mathcal{P}_{[I_{k}, s_{a}, s_{b}]} = \mathcal{I}.
\]

Subsequently, in each local subspace \( \mathcal{P}_{[I_{k}, s_{a}, s_{b}]} \ell_{2}(Z) \) any desired decomposition for property-extraction can be used. When this decomposition is found, we can write for every signal \( x \in \ell_{2}(Z) \):

\[
x = \sum_{k,n} \langle x, \phi_{n}[k] \rangle \phi_{n}[k],
\]

where \( \{ \phi_{n}[k] \mid n = 0, \ldots, \#I_{k} - 1 \} \) is an orthonormal basis in \( \mathcal{P}_{[I_{k}, s_{a}, s_{b}]} \ell_{2}(Z) \), \( k \in Z \). So the next question is how to determine bases for \( \mathcal{P}_{[I_{k}, s_{a}, s_{b}]} \ell_{2}(Z) \). For that we start with defining the symmetrical extension operator \( \mathcal{E}_{[I_{k}, s_{a}, s_{b}]} \), an operator on \( \ell_{2}(I) \) and depends on the discrete-time interval \( I \) and the symmetries \( s_{1}, s_{2} \in \{-1, 1\} \) as parameters.
Definition 10.1.1

The symmetrical extension $E_{[s_1,s_2]}^T$ on $\ell_2([0,L])$ is defined by

\[
E_{[s_1,s_2]}^T(t) = \begin{cases} 
  x(t) & \text{if } t \in [0,L), \\
  s_1 x(-1-t) & \text{if } t \in [-L,0), \\
  s_2 x(2L-1-t) & \text{if } t \in [L, 2L).
\end{cases}
\]  

(10.8)

For arbitrary intervals $I = [a,b]$, we define the symmetrical extension as $E_{[s_1,s_2]}^T = T^{-a}E_{[0,L],s_1,s_2}^T$, where $L = #I = b - a + 1$.

In other words $x$ is extended by the symmetries $s_1$ and $s_2$ resulting in $E_{[s_1,s_2]}^T x$. The numbers $s_1$ and $s_2$ represent symmetries/anti-symmetries with respect to $\frac{-1}{2}$ and $L - \frac{1}{2}$. By the following lemma’s we characterize the local subspace $P_{[l_1,l_2],s_1,s_2}\ell_2(Z)$, in order to compute bases for this subspace.

Lemma 10.1.2

Let $x, y \in P_{[l_1,l_2],s_1,s_2}\ell_2(Z)$, $I = [a,b]$, then there exist $S_x, S_y \in \ell_2(Z)$ such that $x = w[I, \epsilon_{ab}]S_x, y = w[I, \epsilon_{ab}]S_y$ and

\[
\langle x, y \rangle = \langle S_x, S_y \rangle_I = \sum_{t \in I} S_x(t)S_y^*(t).
\]  

Proof: Let $x, y \in P_{[l_1,l_2],s_1,s_2}\ell_2(Z)$, then there exist $g, h \in \ell_2(Z)$ such that $x = P_{[l_1,l_2],s_1,s_2]g$ and $y = P_{[l_1,l_2],s_1,s_2]h$. Define $S_x$ and $S_y$ with support $[a - \epsilon_a, b + \epsilon_b]$, as:

\[
S_x(t) = w(t; I, \epsilon_{ab})g(t) + s_ao w(2a - 1 - t; I, \epsilon_{ab})g(2a - 1 - t) + \\
        s_bw(2b + 1 - t; I, \epsilon_{ab})g(2b + 1 - t),
\]

\[
S_y(t) = w(t; I, \epsilon_{ab})h(t) + s_ao w(2a - 1 - t; I, \epsilon_{ab})h(2a - 1 - t) + \\
        s_bw(2b + 1 - t; I, \epsilon_{ab})h(2b + 1 - t),
\]

for $t \in [a - \epsilon_a, b + \epsilon_b]$. It is clear that $S_x, S_y \in \ell_2(Z)$ and $x = w[I, \epsilon_{ab}]S_x, y = w[I, \epsilon_{ab}]S_y$. We note that $S_x$ and $S_y$ have the symmetries $s_a$ and $s_b$ at the boundaries $a$ and $b$, respectively. Using this property we evaluate

\[
\langle x, y \rangle = \sum_{t=a-\epsilon_a}^{b+\epsilon_b} w^2(t; I, \epsilon_{ab})S_x(t)S_y^*(t)
\]

\[
= \left( \sum_{t=a-\epsilon_a}^{a+\epsilon_a-1} + \sum_{t=a+\epsilon_a}^{b-\epsilon_b} + \sum_{t=b-\epsilon_b+1}^{b+\epsilon_b} \right) w^2(t; I, \epsilon_{ab})S_x(t)S_y^*(t).
\]
Using the symmetrical property of $S_x$ and $S_y$ at $t = a$, we write
\[
\sum_{t=a-\epsilon_a}^{a+\epsilon_a-1} w^2(t; I, \epsilon_{ab}) S_x(t) S_y^*(t) = \sum_{t=a}^{a+\epsilon_a-1} w^2(t; I, \epsilon_{ab}) S_x(t) S_y^*(t) + \sum_{t=a}^{a+\epsilon_a-1} s_a^2 w^2(2a - 1 - t; I, \epsilon_{ab}) S_x(t) S_y^*(t)
\]
\[
= \sum_{t=a}^{a+\epsilon_a-1} S_x(t) S_y^*(t).
\]
Similarly, at $t = b$
\[
\sum_{t=b-\epsilon_a}^{b+\epsilon_a+1} w^2(t; I, \epsilon_{ab}) S_x(t) S_y^*(t) = \sum_{t=b}^{b+\epsilon_a+1} S_x(t) S_y^*(t).
\]
Thus
\[
\langle x, y \rangle = \langle S_x, S_y \rangle_I.
\]

**Lemma 10.1.3** Let $x \in \mathcal{P}_I[I, \epsilon_{1,2}, \epsilon_{1,2}] \ell_2(Z)$, then there exists a vector $p \in \ell_2(I)$ such that $x = w[I, \epsilon_{1,2}] \mathcal{E}[I, \epsilon_{1,2}] p$.

**Proof:** From Lemma 10.1.2, for $x \in \mathcal{P}_I[I, \epsilon_{1,2}, \epsilon_{1,2}] \ell_2(Z)$ there exists $S_x \in \ell_2(Z)$ such that $x = w[I, \epsilon_{1,2}] S_x$. Define $p \in \ell_2(I)$ on $I$ by
\[
p(t) = \begin{cases} 
    w(t; I, \epsilon_{1,2}) x(t) + s_1 w(-1 - t; I, \epsilon_{1,2}) x(-1 - t) & \text{if } t \in [0, \epsilon_1 - 1], \\
    x(t) & \text{if } t \in [\epsilon_1, L-\epsilon_2 - 1], \\
    w(t; I, \epsilon_{1,2}) x(t) + s_2 w(2L - 1 - t; I, \epsilon_{1,2}) x(2L - 1 - t) & \text{if } t \in [L - \epsilon_2, L - 1]. 
\end{cases}
\]
Consequently,
\[
x[-\epsilon_1, L+\epsilon_2 - 1] S_x = x[-\epsilon_1, L+\epsilon_2 - 1] \mathcal{E}[I, \epsilon_{1,2}] p.
\]

**Lemma 10.1.4** Let $x, y \in \ell_2(I)$, then
\[
\langle w[I, \epsilon_{1,2}] \mathcal{E}[I, \epsilon_{1,2}] x, w[I, \epsilon_{1,2}] \mathcal{E}[I, \epsilon_{1,2}] y \rangle = \langle x, y \rangle.
\]
\[\text{(10.9)}\]
10.1 Lapped orthogonal segmentation for discrete signals

Proof:

\[
\langle w[I, \epsilon_{1,2}] \xi_{[I, \epsilon_{1,2}]} x, w[I, \epsilon_{1,2}] \xi_{[I, \epsilon_{1,2}]} y \rangle = \\
= \sum_{-\epsilon_1 \leq t \leq -1} w^2(t; I, \epsilon_1, \epsilon_2)(\xi_{[I, \epsilon_{1,2}]} x)(t)(\xi_{[I, \epsilon_{1,2}]} y)^*(t) \\
+ \sum_{\epsilon_1 \leq t \leq L-\epsilon_2-1} w^2(t; I, \epsilon_1, \epsilon_2)(\xi_{[I, \epsilon_{1,2}]} x)(t)(\xi_{[I, \epsilon_{1,2}]} y)^*(t) \\
+ \sum_{L-\epsilon_2 \leq t \leq L+\epsilon_2-1} w^2(t; I, \epsilon_1, \epsilon_2)(\xi_{[I, \epsilon_{1,2}]} x)(t)(\xi_{[I, \epsilon_{1,2}]} y)^*(t),
\]

where

\[
\sum_{-\epsilon_1 \leq t \leq -1} w^2(t; I, \epsilon_1, \epsilon_2)(\xi_{[I, \epsilon_{1,2}]} x)(t)(\xi_{[I, \epsilon_{1,2}]} y)^*(t) = \\
= \sum_{-\epsilon_1 \leq t \leq -1} w^2(t; I, \epsilon_1, \epsilon_2)(\xi_{[I, \epsilon_{1,2}]} x)(t)(\xi_{[I, \epsilon_{1,2}]} y)^*(t) + \\
\sum_{0 \leq t \leq \epsilon_2-1} w^2(t; I, \epsilon_1, \epsilon_2)(\xi_{[I, \epsilon_{1,2}]} x)(t)(\xi_{[I, \epsilon_{1,2}]} y)^*(t) \\
= \sum_{-\epsilon_1 \leq t \leq -1} s^2 w^2(t; I, \epsilon_1, \epsilon_2)x(-1-t)y^*(-1-t) + \\
\sum_{0 \leq t \leq \epsilon_2-1} w^2(t; I, \epsilon_1, \epsilon_2)x(t)y^*(t) \\
= \sum_{0 \leq t \leq \epsilon_2-1} (w^2(t; I, \epsilon_1, \epsilon_2) + w^2(-1-t; I, \epsilon_1, \epsilon_2))x(t)y^*(t) \\
= \sum_{0 \leq t \leq \epsilon_2-1} x(t)y^*(t).
\]  
(10.10)

Similarly,

\[
\sum_{L-\epsilon_2 \leq t \leq L+\epsilon_2-1} (\xi_{[I, \epsilon_{1,2}]} x)(t)(\xi_{[I, \epsilon_{1,2}]} y)^*(t) = \sum_{L-\epsilon_2 \leq t \leq L-1} x(t)y^*(t).
\]

Further,

\[
\sum_{\epsilon_1 \leq t \leq L-\epsilon_2-1} w^2(t; I, \epsilon_1, \epsilon_2)(\xi_{[I, \epsilon_{1,2}]} x)(t)(\xi_{[I, \epsilon_{1,2}]} y)^*(t) = \sum_{\epsilon_1 \leq t \leq L-\epsilon_2-1} x(t)y^*(t).
\]

In the following theorem we summarize the properties of the local subspace obtained by applying the orthogonal projection \(P_{[I, \epsilon_{1,2}]}\) to \(\ell_2(Z)\).
Theorem 10.1.5  The operator defined by $x \mapsto w[I, \epsilon_{1,2}]E_{[I, \alpha_1, \alpha_2]}x$ is a unitary map from $\ell_2(I)$ onto the subspace $\mathcal{P}_{[I, \alpha_1, \alpha_2]} \ell_2(\mathbb{Z})$.  

Let $\{u_n|n=0, \ldots, L-1\}$ be an orthonormal basis in $\ell_2(I), I = [a, b], L = \#I$, and extend the support $I$ of this basis over $[a - 2L, b + L]$ by the orthogonal system $\{E_{[I, \nu, \nu]}u_n|n=0, \ldots, L-1\}$ in $\ell_2([a - 2L, b + L])$. Let $w[I, \epsilon_{ab}]$ be a window function satisfying, see (10.3), the power-complementarity condition

$$w^2(t; I, \epsilon_{ab}) + w^2(2a - 1 - t; I, \epsilon_{ab}) + w^2(2b + 1 - t; I, \epsilon_{ab}) = 1,$$

for $t \in [a - \epsilon_a, b + \epsilon_b]$. Then in the following theorem an orthonormal basis for $\mathcal{P}_{[I, \epsilon_{abo}, \epsilon_{a'b'}]} \ell_2(\mathbb{Z})$ is presented.

Theorem 10.1.6  Let $\{u_n|n=0, \ldots, L-1\}$ be an orthonormal basis in $\ell_2(I)$. Then the system $\{w[I, \epsilon_{ab}]E_{[I, \nu, \nu]}u_n|n=0, \ldots, L-1\}$

is an orthonormal basis of $\mathcal{P}_{[I, \epsilon_{abo}, \epsilon_{a'b'}]} \ell_2(\mathbb{Z})$.

Proof:  Let $x \in \mathcal{P}_{[I, \epsilon_{abo}, \epsilon_{a'b'}]} \ell_2(\mathbb{Z})$, by Lemma 10.1.3, there exists $p \in \ell_2(I)$ such that

$$x = w[I, \epsilon_{ab}]E_{[I, \nu, \nu]}p.$$ 

Since $p \in \ell_2(I)$ we write

$$p = \sum_{n=0}^{L-1} \langle p, u_n \rangle u_n.$$ 

From Lemma 10.1.4, we have

$$\langle p, u_n \rangle = \langle w[I, \epsilon_{ab}]E_{[I, \nu, \nu]}p, w[I, \epsilon_{ab}]E_{[I, \nu, \nu]}u_n \rangle = \langle x, w[I, \epsilon_{ab}]E_{[I, \nu, \nu]}u_n \rangle.$$ 

Thus,

$$x = \sum_{n=0}^{L-1} \langle x, w[I, \epsilon_{ab}]E_{[I, \nu, \nu]}u_n \rangle w[I, \epsilon_{ab}]E_{[I, \nu, \nu]}u_n.$$ 

Definition 10.1.7  The orthonormal basis $\{w[I, \epsilon_{ab}]E_{[I, \nu, \nu]}u_n|n=0, \ldots, L-1\}$ is called a lapped orthonormal basis for $\mathcal{P}_{[I, \epsilon_{abo}, \epsilon_{a'b'}]} \ell_2(\mathbb{Z})$. 

In the next theorem we present a basis for the segmented signal space $\ell_2(\mathbb{Z})$. 

168  Orthogonal transformations by 1-D domain segmentations
10.1 Lapped orthogonal segmentation for discrete signals

**Theorem 10.1.8** Let \( \{I_k|k \in \mathbb{Z}\} \) be a segmentation of \( \mathbb{Z} \), \( I_k = [a_k, a_{k+1}) \). Further, let \( \{u_{n,k}|n = 0, \ldots, L_k - 1\} \) be an orthonormal basis in \( \ell_2(I_k), L_k = \#I_k, k \in \mathbb{Z} \). Then the system

\[
\{w(t; I_k, \epsilon_k) \langle E_{[I,s_k]}u_{n,k} \rangle (t - a_k)|n = 0, \ldots, L_k - 1, k \in \mathbb{Z}\}
\]

(10.12)

is an orthonormal basis of \( \ell_2(\mathbb{Z}) \). Further, we know that \( s_{a_k} = -s_{-1+a_k} \), \( \epsilon_{a_k} = \epsilon_{-1+a_k} \).

**Proof:** From Theorem 10.1.6, for each \( \epsilon_{\omega} \in \mathbb{Z} \), the system

\[
\{w(t; I_k, \epsilon_k) \langle E_{[I,s_k]}u_{n,k} \rangle (t - a_k)|n = 0, \ldots, L_k - 1\}
\]

constitutes an orthonormal basis for \( \mathcal{P}[I_{k,\epsilon,\omega}]\ell_2(\mathbb{Z}) \). Further, we know that \( \ell_2(\mathbb{Z}) = \theta_{k \in \mathbb{Z}} \mathcal{P}[I_{k,\epsilon,\omega}]\ell_2(\mathbb{Z}) \), since \( s_{a_k} = -s_{-1+a_k} \) and \( \epsilon_{a_k} = \epsilon_{-1+a_k} \). Thus the system

\[
\{w(t; I_k, \epsilon_k) \langle E_{[I,s_k]}u_{n,k} \rangle (t - a_k)|n = 0, \ldots, L_k - 1, k \in \mathbb{Z}\}
\]

constitutes an orthonormal basis for \( \ell_2(\mathbb{Z}) \). \( \square \)

From the famous Fourier series, four bases with different symmetrical properties at the interval boundaries can be constructed, namely two sine and two cosine bases. In the following theorem we present these bases and give their corresponding symmetries at the boundaries.

**Theorem 10.1.9** Let \( I = [a, b) \) be a discrete-time interval, \( L = \#I = b - a \). In the following we present bases with their corresponding local subspaces in \( \ell_2(\mathbb{Z}) \):

1. The system \( \{w[I, \epsilon_{\omega}] \sqrt{\frac{2}{L}} \sin \left( \frac{\pi}{L}(n + \frac{1}{2})(t + \frac{1}{2} - a) \right) |n = 0, \ldots, L - 1 \} \) constitutes a local orthonormal basis in \( \mathcal{P}[I_{\epsilon,\omega}, -1, -\frac{1}{2}]\ell_2(\mathbb{Z}) \).

2. The system \( \{w[I, \epsilon_{\omega}] \sqrt{\frac{2}{L}} \sin \left( \frac{\pi}{L}(n(t + \frac{1}{2} - a)) |n = 0, \ldots, L - 1 \} \) constitutes a local orthonormal basis in \( \mathcal{P}[I_{\epsilon,\omega}, -1, -\frac{1}{2}]\ell_2(\mathbb{Z}) \).

3. The system \( \{w[I, \epsilon_{\omega}] \sqrt{\frac{2}{L}} \cos \left( \frac{\pi}{L}(n + \frac{1}{2})(t + \frac{1}{2} - a) \right) |n = 0, \ldots, L - 1 \} \) constitutes a local orthonormal basis in \( \mathcal{P}[I_{\epsilon,\omega}, 1, -\frac{1}{2}]\ell_2(\mathbb{Z}) \).

4. The system \( \{w[I, \epsilon_{\omega}] \sqrt{\frac{2}{L}} \cos \left( \frac{\pi}{L}(n(t + \frac{1}{2} - a)) |n = 1, \ldots, L - 1 \} \cup \{w[I, \epsilon_{\omega}] \sqrt{\frac{2}{L}} \cos \left( \frac{\pi}{L}(n + \frac{1}{2})(t + \frac{1}{2} - a) \right) |n = 0, \ldots, L - 1 \} \) constitutes a local orthonormal basis in \( \mathcal{P}[I_{\epsilon,\omega}, 1, -\frac{1}{2}]\ell_2(\mathbb{Z}) \).
Example 10.1.10 We return to the illustration in the introduction, namely the uttered sentence "You’re cool!". Here we segmented the signal by lapped orthogonal projections, and decompose it by local bases. We segmented $I$ into 5 equidistant intervals. We used the cosine basis for the decomposition, the third basis in Theorem 10.1.9. The symmetries are $(-1,1)$ for each interval. Depending on the energy distribution we have chosen 984 largest coefficient, containing 99.4% of the energy of $x$. In Fig. 10.1, we show plots of an approximation for $x$, its Fourier spectrum and the loss between the original $x$ and its approximation. The plot in Fig. 10.2 shows the local Fourier spectrum of the approximation of $x$. We observe that the artifacts are completely absent in the local spectrum. Further, in comparison with the approximations in the introduction, the approximation based on lapped segmentation compresses more accurately the signal.

![Figure 10.1: Signal approximation, frequency spectrum and the difference between the original signal and its reconstruction in the case of smooth orthogonal projections.](image)

Intuitively, it is clear that the local bases described above, especially the trigonometric bases, do not fully cover the time-frequency plane, since the frequencies in the bases form an equidistant countable set in the continuous-frequency interval, per time interval. This means to approximate $w[I, \epsilon_1, 2] \cos(\omega t)$, (with frequency $\omega$ not belonging to the set of frequencies identified by the basis function, bounded by the set), by the cosine basis in $\mathcal{P}_{I,\epsilon_1,2,s_1,*,s_2}\ell_2(\mathbb{Z})$, many coefficients are needed to achieve an accurate approximation. To make the goal of this discussion clear, it were better if we would have more freedom in choosing the frequencies in the bases, in order to approximate signals accurately depending on their frequency character, without loosing the (local) orthogonal segmentation. At this stage, two questions arise. First,
how to determine these frequencies, and second, what are the systems identified by these frequencies, for example, bases / complete systems. The first question comes down to determining peaks in the local spectra. In the following we discuss how this freedom can be interpreted. Completeness is a necessary condition to represent signals in subspaces of $\ell_2(\mathbb{Z})$. Further, since in signal processing we deal with analysis and synthesis, simple representation schemes for signals by these complete systems are required. Frames contain the freedom as described above. Frame theory guarantees the completeness, stability and redundancy of linear signal representation. Let shortly recall the main aspects of frames. A frame is a complete system that characterize signals from their inner products with the frame elements. Let $\{\phi_n|n \in \Gamma\}$ be a frame in the Hilbert space $\mathcal{H}$, which is in this case finite dimensional, and let $\Phi : \mathcal{H} \longrightarrow \ell_2(\Gamma)$ be the frame operator defined by

$$ (\Phi x)(n) = \langle x, \phi_n \rangle, \quad x \in \mathcal{H}, n \in \Gamma, \quad (10.13) $$

then there exists a system $\{\psi_n|n \in \Gamma\}$, such that

$$ x = \sum_{n \in \Gamma} \langle x, \psi_n \rangle \phi_n = \sum_{n \in \Gamma} \langle x, \phi_n \rangle \psi_n. \quad (10.14) $$

Such a system $\{\psi_n|n \in \Gamma\}$ is the dual frame of $\{\phi_n|n \in \Gamma\}$, defined by

$$ \psi_n := (\Phi^* \Phi)^{-1} \phi_n, $$
172 Orthogonal transformations by 1-D domain segmentations

for \( n \in \Gamma \). The frame bounds, in terms of the frame operator are given by

\[
A = \| (\Phi^* \Phi)^{-1} \|^{-1}, \\
B = \| \Phi^* \Phi \| = \| \Phi \|^2.
\]

The ratio \( A/B \) indicates how conditionally stable the frame is. In general, we are more interested in conditionally stable frames \( A/B \approx 1 \). The frame is tight if \( A = B \).

The operator \( \Phi \Phi^* \) is described with the transpose of the Gram-matrix \( G, [\Phi \Phi^*] = G^\top \), obtained by

\[
G(i, j) = \langle \psi_i, \phi_j \rangle, \ i, j \in \Gamma.
\]  

(10.15)

For local bases we have seen how to construct a basis for \( \ell_2(\mathbb{Z}) \) from a basis in \( \ell_2(I) \). We repeat this question for frames. The answer to this question is positive, in the sense that each frame in \( \ell_2(\mathbb{Z}) \) yields a frame in \( \ell_2(I) \).

**Theorem 10.1.11** Let \( \{ \phi_n | n \in \Gamma \} \) be a frame in \( \ell_2(I) \), \( \# \Gamma \geq \# I \) with frame bounds \( A, B \) and dual frame \( \{ \psi_n | n \in \Gamma \} \). Then the system

\[
\{ w[I, \epsilon_{1,2}] \mathcal{E}_{[I, \sigma_1, \sigma_2]} \phi_n | n \in \Gamma \}
\]

(10.16)

is a lapped frame in \( \mathcal{P}_{[I, \epsilon_{1,2}, \sigma_1, \sigma_2]} \ell_2(\mathbb{Z}) \) with the same frame bounds \( A, B \) and dual frame \( \{ w[I, \epsilon_{1,2}] \mathcal{E}_{[I, \sigma_1, \sigma_2]} \psi_n | n \in \Gamma \} \).

**Proof:** Let \( y \in \mathcal{P}_{[I, \epsilon_{1,2}, \sigma_1, \sigma_2]} \ell_2(\mathbb{Z}) \), then by Lemma 10.1.3 there exists \( p \in \ell_2(I) \) such that

\[
y = w[I, \epsilon_{1,2}] \mathcal{E}_{[I, \sigma_1, \sigma_2]} p,
\]

where by Lemma 10.1.4

\[
\langle y, y \rangle = \langle p, p \rangle \\
\langle y, w[I, \epsilon_{1,2}] \mathcal{E}_{[I, \sigma_1, \sigma_2]} \phi_n \rangle = \langle p, \phi_n \rangle.
\]

Since \( \{ \phi_n | n \in \Gamma \} \) is a frame in \( \ell_2(I) \), there exist \( A, B > 0 \) such that for all \( x \in \ell_2(I) \)

\[
A \| x \|^2 \leq \sum_{n \in \Gamma} | \langle x, \phi_n \rangle |^2 \leq B \| x \|^2.
\]

Thus,

\[
A \| y \|^2 \leq \sum_{n \in \Gamma} \| \langle y, w[I, \epsilon_{1,2}] \mathcal{E}_{[I, \sigma_1, \sigma_2]} \phi_n \rangle \|^2 \leq B \| y \|^2.
\]
For $p \in \ell_2(I)$ we know that

$$p = \sum_{n \in \Gamma} \langle p, \psi_n \rangle \phi_n$$

Equivalently, applying the unitary map $x \mapsto w[I, \epsilon_{1,2}] \mathcal{E}_{[I, s_1, s_2]} x$ we have

$$y = \sum_{n \in \Gamma} \langle y, w[I, \epsilon_{1,2}] \mathcal{E}_{[I, s_1, s_2]} \psi_n \rangle w[I, \epsilon_{1,2}] \mathcal{E}_{[I, s_1, s_2]} \phi_n.$$ 

Representations based on tight frames seem to have all nice properties of representations based on orthonormal bases. But there is the additional advantage, if we apply any orthogonal projection to any tight frame then we obtain a tight frame for the closed subspace equal to the range of the orthogonal projection. The next theorem tells us how simple it can be to construct tight frames for any subspace of signals.

**Theorem 10.1.12 (van Eijndhoven)** Let the collection $\{\phi_n | n \in \Gamma\}$ be in the Hilbert space $\mathcal{H}$. Then $\{\phi_n | n \in \Gamma\}$ is a tight frame with frame bound one if and only if there exists a Hilbert space $\mathcal{H}_1$ and an orthonormal basis $\{\psi_n | n \in \Gamma\}$ in the direct Hilbert sum $\mathcal{H} \oplus \mathcal{H}_1$, such that for all $n \in \Gamma$

$$\phi_n = \mathcal{P} \psi_n,$$

where $\mathcal{P}$ denotes the orthogonal projection of $\mathcal{H} \oplus \mathcal{H}_1$ onto $\mathcal{H}$. 

**Proof:** Assume that there exists a Hilbert space $\mathcal{H}_1$ and an orthonormal basis $\{\psi_n | n \in \Gamma\}$ in the direct Hilbert sum $\mathcal{H} \oplus \mathcal{H}_1$, such that for all $n \in \Gamma$

$$\phi_n = \mathcal{P} \psi_n,$$

where $\mathcal{P}$ denotes the orthogonal projection of $\mathcal{H} \oplus \mathcal{H}_1$ onto $\mathcal{H}$. Then for $x \in \mathcal{H}$:

$$\|x\|^2 = \sum_{n \in \Gamma} |\langle x, \psi_n \rangle|^2 = \sum_{n \in \Gamma} |\langle \mathcal{P} x, \psi_n \rangle|^2 = \sum_{n \in \Gamma} |\langle x, \phi_n \rangle|^2.$$ 

Hence $\{\phi_n | n \in \Gamma\}$ is a tight frame in $\mathcal{H}$ with frame bound one.

Let $\{\phi_n | n \in \Gamma\}$ be a tight frame in $\mathcal{H}$ with frame bound, and let $\mathcal{A}$ be the frame operator. Then $\mathcal{A}^* \mathcal{A} = \mathcal{I}$, that is, $\mathcal{A}$ is an isometry and consequently $\mathcal{P} = \mathcal{A} \mathcal{A}^*$ is the orthogonal projection onto range($\mathcal{A}$), which is a closed subspace in $\ell_2(\Gamma')$. Let $\mathcal{H}_1$ be ker($\mathcal{A}^*$), then $\mathcal{H}_1$ is a Hilbert space being a closed subspace in $\ell_2(\Gamma')$. Consider the direct sum Hilbert space:

$$\tilde{\mathcal{H}} = \mathcal{H} \oplus \mathcal{H}_1,$$
which consists of all ordered pairs \( x \oplus y \), with \( x \in \mathcal{H} \) and \( y \in \mathcal{H}_1 \), and is endowed with the inner product
\[
\langle x_1 \oplus y_1, x_2 \oplus y_2 \rangle = \langle x_1, x_2 \rangle + \langle y_1, y_2 \rangle.
\]
Since \( \mathcal{H}_1 = \ker(\mathcal{A}^*) = \text{range}(\mathcal{A})^\perp \), \( \mathcal{I} - \mathcal{P} \) is the orthogonal projection onto \( \ker(\mathcal{A}^*) \).
Define \( \psi_n \in \mathcal{H} \) by
\[
\psi_n = \phi_n \oplus (\mathcal{I} - \mathcal{P})e_n = \mathcal{A}^* e_n \oplus (\mathcal{I} - \mathcal{P})e_n.
\]
Then
\[
\langle \psi_n, \psi_m \rangle = \langle \phi_n, \phi_m \rangle + \langle (\mathcal{I} - \mathcal{P})e_n, (\mathcal{I} - \mathcal{P})e_m \rangle = \langle \mathcal{P} e_n, e_m \rangle + \langle (\mathcal{I} - \mathcal{P})e_n, e_m \rangle = \langle e_n, e_m \rangle = \delta_{nm}.
\]
Hence \( \{\psi_n | n \in \Gamma\} \) is an orthonormal system in \( \tilde{\mathcal{H}} \). Further, for \( x \in \mathcal{H} \) and \( y \in \ell_2(\Gamma) \)
\[
\sum_{n \in \Gamma} \langle x \oplus (\mathcal{I} - \mathcal{P})y, \psi_n \rangle = \sum_{n \in \Gamma} \langle Ax \oplus (\mathcal{I} - \mathcal{P})y, e_n \rangle (\mathcal{A}^* e_n \oplus (\mathcal{I} - \mathcal{P})e_n)
= \mathcal{A}^* (Ax \oplus (\mathcal{I} - \mathcal{P})y) \oplus (\mathcal{I} - \mathcal{P}) (Ax \oplus (\mathcal{I} - \mathcal{P})y)
= x \oplus (\mathcal{I} - \mathcal{P})y.
\]
It follows that \( \{\phi_n | n \in \Gamma\} \) is an orthonormal basis in \( \tilde{\mathcal{H}} \), where
\[
\phi_n = \mathcal{P} \psi_n,
\]
for all \( n \in \Gamma \).

This theorem is formulated in the general context of an abstract Hilbert space. In the one-dimensional case, tight frames can be constructed as follows. Let \( U \) be any orthogonal \( K \times K \)-matrix, \( U U^T = I_K, K \geq L \). Divide the matrix \( U \) into two parts \( U_1 \) and \( U_2 \) such that
\[
U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix},
\]
where \( U_1 \) is \( L \times K \)-matrix. Consequently, \( U_1 U_1^T = I_L \), which means that the system obtained by the columns of \( U_1 \) constitutes a tight frame in \( \mathbb{C}^L \). Next we give an example of tight frames obtained by the cosine basis. For a tight frame \( \{\phi_n | n \in \Gamma\} \) in \( \ell_2(I) \), there exists \( A > 0 \) such that
\[
\|x\|^2 = A \sum_{n \in \Gamma} |\langle x, \phi_n \rangle|^2,
\]
for all \( x \in \ell_2(I) \). This property is similar to the Parseval equality, which means that using tight frames in local segmentation also preserves energy, by a scale factor.
Lapped orthogonal segmentation for continuous signals

Example 10.1.13 The system \( \{ \sqrt{\frac{2}{K}} \cos \left( \frac{\pi}{K} (n + \frac{1}{2}) (t + \frac{1}{2}) \right) \mid n = 0, \ldots, K - 1 \} \) constitutes an orthonormal basis for \( \ell_2([0,K]) \). Let \( U \) be the \( K \times K \)-matrix constructed from the elements of the orthonormal system \( \{ \sqrt{\frac{2}{K}} \cos \left( \frac{\pi}{K} (n + \frac{1}{2}) (t + \frac{1}{2}) \right) \mid n = 0, \ldots, K - 1 \} \) as columns. Let \( L \in \mathbb{N}, L \leq K \) and divide the matrix \( U \) into \( U_1 \) and \( U_2 \) such that

\[
U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix},
\]

where \( U_1 \) is an \( L \times K \)-matrix. Then the system \( \{ \sqrt{\frac{2}{K}} \cos \left( \frac{\pi}{K} (n + \frac{1}{2}) (t + \frac{1}{2} - a) \right) \mid n = 0, \ldots, K - 1, t \in I \} \) constitutes a tight frame in \( \ell_2(I) \), \( I = [a, b), L = \# I \). Consequently, the system \( \{ u[I,e_{ab}] \sqrt{\frac{2}{K}} \cos \left( \frac{\pi}{K} (n + \frac{1}{2}) (t + \frac{1}{2} - a) \right) \mid n = 0, \ldots, K - 1 \} \) constitutes a local frame in \( P_{[I,e_{ab},1,-1]} \ell_2(\mathbb{Z}) \). We see that this system contains more frequency choice than the local basis. This local frame can be used to approximate the frequencies more accurately. \( \square \)

10.2 Lapped orthogonal segmentation for continuous signals

The discrete-time lapped orthogonal segmentation can be seen as a uniformly sampled version of the continuous-time lapped orthogonal segmentation. The main difference between the two cases are the points of symmetries (cutoff). In discrete-time case, the points of symmetry are the half-integers between the boundaries of adjacent intervals. In continuous-time case, they are the boundaries themselves. In the literature one starts with the continuous-time situations, and then derives similar results for the discrete-time situations, cf. [3, 64]. From a practical (numerical) point of view and avoiding the confusion in the points of symmetries, we have chosen for a discrete-time start. Nevertheless, the continuous-time case contains the beauty in construction and derivation of the results, even more, it has the scaling property, for example constructing a basis in \( L_2(a,b) \) from a basis in \( L_2(0,1) \). This property is missing in the discrete-time case. This way, it is sufficient to construct bases for \( L_2(0,1) \), stretch it to different intervals \([a,b]\) and then sample it to discrete-time intervals. The results in this section can be proved in a similar way as in the previous section for discrete-time case. Let us start from the orthogonal projections \( P_{[\mathbb{R}_0^+,e,s^+]} \), \( P_{[\mathbb{R}_0^-,e,s^-]} \) identified by \( \mathbb{R}_0^+ \) and \( \mathbb{R}_0^- \):

\[
\begin{align*}
(P_{[\mathbb{R}_0^+,e,s^+]}) (t) &= u^2(t;e)x(t) + s^+ w(t;e) w(-t;e) x(-t), \ t \in \mathbb{R} \\
(P_{[\mathbb{R}_0^-,e,s^-]}) (t) &= u^2(-t;e)x(t) + s^- w(-t;e) w(t;e) x(-t), \ t \in \mathbb{R}
\end{align*}
\]
where the window function \( w[e] \) is satisfying the power-complementarity condition

\[
\begin{align*}
w^2(t; e) + w^2(-t; e) &= 1,
\end{align*}
\]

support\( (w[e]) = [-\epsilon, \infty), \)

and

\[
\begin{align*}
s^- = -s^+ \in \{-1, 1\},
\end{align*}
\]

guaranteeing that \( \mathcal{P}_{[\mathbb{R}_0^+, e, s^+] + \mathcal{P}_{[\mathbb{R}^-, e, s^-]} = I \) and therefore \( \mathcal{P}_{[\mathbb{R}_0^+, e, s^+] \mathcal{P}_{[\mathbb{R}^-, e, s^-]} = 0. \)

It is clear that \( w(t; e) = 1, \) for \( t \geq \epsilon. \) We note that if \( \epsilon = 0, \) then \( w[0] = \chi_{\mathbb{R}_0^+}, \)

and the orthogonal projections \( \mathcal{P}_{[\mathbb{R}_0^+, 0, s^+]} \mathcal{P}_{[\mathbb{R}^-, 0, s^-]} \) become multiplication by the corresponding character functions. The window function \( w[e] \) guarantees the smooth transition in the cutoff \( 0. \)

**Example 10.2.1** From the power-complementarity condition, the function \( w[e] \) can be constructed as \( w(t; e) = \sin \theta(t; e), \) where \( \theta \) is smoothly increasing from \( 0 \) to \( \pi/2 \) in \( [-\epsilon, \epsilon], \) for instance

\[
\begin{align*}
\theta(t; e) &= \int_{-\infty}^{t} \psi(\tau; e) d\tau,
\end{align*}
\]

where \( \psi[e] \) is even non-negative, support\( (\psi[e]) \subseteq [-\epsilon, \epsilon], \) and

\[
\begin{align*}
\int_{-\infty}^{\infty} \psi(\tau; e) d\tau &= \frac{\pi}{2},
\end{align*}
\]

Consequently, we can write for \( \theta(t; e) \)

\[
\begin{align*}
\theta(t; e) + \theta(-t; e) &= \frac{\pi}{2},
\end{align*}
\]

which means for \( w(-t; e) \)

\[
\begin{align*}
w(-t; e) = \sin \theta(-t; e)
&= \cos \theta(t; e).
\end{align*}
\]

\[\square\]

For a real interval \( I = [a, b], a + \epsilon_a \leq b - \epsilon_b, \) the orthogonal projection \( \mathcal{P}_{[I, e_a, s_a, s_b]}, \)

\( s_a, s_b \in \{-1, 1\}, e_a, e_b \in \mathbb{R}_0^+ \times \mathbb{R}_0^+, \) is defined by:

\[
\begin{align*}
\mathcal{P}_{[I, e_a, s_a, s_b]} = T^{-a} \mathcal{P}_{[\mathbb{R}_0^+, e_a, s_a]} T^{a-b} \mathcal{P}_{[\mathbb{R}_0^-, e_a, s_a]} T^b,
\end{align*}
\]
Lapped orthogonal segmentation for continuous signals

where \( T^{a-b} = T^{-b}T^{a} \). The definition of \( P_{[I\notin\{s_{a},s_{b}\}]} \) makes sense because the projections \( T^{-a}P_{[\tilde{R}^{+}\notin\{s_{a}\}]}T^{a} \) and \( T^{-b}P_{[\tilde{R}^{-}\notin\{s_{b}\}]}T^{b} \) are orthogonal and

\[
T^{-a}P_{[\tilde{R}^{+}\notin\{s_{a}\}]}T^{a} = T^{-b}P_{[\tilde{R}^{-}\notin\{s_{b}\}]}T^{b} = T^{-a}P_{[\tilde{R}^{-}\notin\{s_{b}\}]}T^{a} = T^{a-b}P_{[I\notin\{s_{a},s_{b}\}]}T^{a}.
\]

Explicitly,

\[
\left(P_{[I\notin\{s_{a},s_{b}\}]}x\right)(t) = w(t; I, \epsilon_{ab}) \left[ w(t; I, \epsilon_{ab})x(t) + s_{a}w(2a - t; I, \epsilon_{ab})x(2a - t) + s_{b}w(2b - t; I, \epsilon_{ab})x(2b - t) \right]
\]

(10.19)

where \( w(t; I, \epsilon_{ab}) \) can be written in terms of \( w[t] \), as follows:

\[
w(t; I, \epsilon_{ab}) = w(t - \epsilon_{b}a; \epsilon_{a})w(b - t; \epsilon_{b}), \quad a - \epsilon_{a} \leq t \leq b + \epsilon_{b},
\]

(10.20)

satisfying the power-complementarity condition:

\[
w^{2}(t; I, \epsilon_{ab}) + w^{2}(2a - t; I, \epsilon_{ab}) + w^{2}(2b - t; I, \epsilon_{ab}) = 1,
\]

(10.21)

for all \( t \in \text{support}(w[I, \epsilon_{ab}]) = [a - \epsilon_{a}, b + \epsilon_{b}] \). Similarly to the discrete-time case we can write for two adjacent intervals:

\[
w^{2}(t; I, \epsilon_{ab}) + w^{2}(t; J, \epsilon_{b,c}) = w^{2}(t - \epsilon_{a}; \epsilon_{b})w^{2}(c - t; \epsilon_{c}) = w^{2}(t; I \cup J, \epsilon_{ac}),
\]

(10.22)

\[
P_{[I\notin\{s_{a},s_{b}\}]} + P_{[J\notin\{s_{b},s_{c}\}]} = P_{[I \cup J \notin\{s_{b},s_{c}\}]}.
\]

(10.23)

In this way, given a partition \( \{I_{k}\}_{k \in \mathbb{Z}} \) of the real axis, the projections \( P_{[I_{k}]} \), \( \epsilon_{k} = (\epsilon_{k}, \epsilon_{k+1}) \in \mathbb{R}_{+}^{2} \times \mathbb{R}_{+}^{2}, s_{k+1}(1) = -s_{k}(2), s_{k} \in \{-1, 1\} \times \{-1, 1\}, \) decompose the signal space \( L_{2}(\mathbb{R}) \) into a direct sum of mutually orthogonal subspaces, that is, they form a resolution of the identity. In other words, there is the following decomposition:

\[
L_{2}(\mathbb{R}) = \oplus_{k \in \mathbb{Z}} P_{[I_{k}]}L_{2}(\mathbb{R})
\]

that is

\[
\sum_{k \in \mathbb{Z}} P_{[I_{k}]} = I.
\]

For local bases in \( P_{[I]}L_{2}(\mathbb{R}) \), \( I = [a, b] \), we can derive similar results to those of the discrete-time case. Moreover, in the continuous-time case it is sufficient to start with a basis in \( L_{2}(0, 1) \), that is, the system \( \{\phi_{n}(t)\}_{n \in \mathbb{N}_{0}} \) constitutes an
orthonormal basis in \(L_2(0, 1)\) iff \(\{\frac{1}{\sqrt{b-a}}\phi_n(t-a)\mid n \in \mathbb{N}_0\}\) constitutes an orthonormal basis in \(L_2(a, b)\). We start with defining the symmetrical extension operator \(E_{[I, s_a, s_b]}\) on \(L_2(I)\), for instance \(I = [0, 1]\), by:

\[
(E_{[I, s_a, s_b]}x)(t) = \begin{cases} 
 x(t) & \text{if } t \in [0, 1], \\
 s_0 x(-t) & \text{if } t \in [-1, 0), \\
 s_1 x(2-t) & \text{if } t \in (1, 2]. 
\end{cases}
\]

**Lemma 10.2.2** Let \(x, y \in P_{[I, \epsilon_\alpha, \epsilon_\beta, \epsilon_\gamma]}L_2(\mathbb{R})\), then there exist \(S_x, S_y \in L_2(\mathbb{R})\) such that \(x = w[I, \epsilon_\alpha]S_x, y = w[I, \epsilon_\alpha]S_y\) and

\[
\langle x, y \rangle = \langle S_x, S_y \rangle_I = \int_a^b S_x(t)S_y^*(t)dt.
\]

**Lemma 10.2.3** Let \(x \in P_{[I, \epsilon_\alpha, \epsilon_\beta, \epsilon_\gamma]}L_2(\mathbb{R})\), then there exists a vector \(p \in L_2(I)\) such that \(x = w[I, \epsilon_\alpha]\epsilon_{[I, s_a, s_b]}p\).

**Lemma 10.2.4** Let \(x, y \in L_2(I)\), then

\[
\langle w[I, \epsilon_\alpha]\epsilon_{[I, s_a, s_b]}x, w[I, \epsilon_\alpha]\epsilon_{[I, s_a, s_b]}y \rangle = \langle x, y \rangle. \tag{10.24}
\]

The characterization of the local subspace \(P_{[I, \epsilon_\alpha, \epsilon_\beta, \epsilon_\gamma]}L_2(\mathbb{R})\) is given in the following theorem.

**Theorem 10.2.5** The operator defined by \(\phi \mapsto w[I, \epsilon_\alpha]\epsilon_{[I, s_a, s_b]}\phi\) is a unitary map from \(L_2(I)\) onto the subspace \(P_{[I, \epsilon_\alpha, \epsilon_\beta, \epsilon_\gamma]}L_2(\mathbb{R})\).

Similarly to the discrete-time case the following theorem present an orthonormal basis for \(P_{[I, \epsilon_\alpha, \epsilon_\beta, \epsilon_\gamma]}L_2(\mathbb{R})\).

**Theorem 10.2.6** Let \(\{u_n\mid n \in \mathbb{N}_0\}\) be an orthonormal basis in \(L_2(I)\). Then the system

\[
\{w[I, \epsilon_\alpha]\epsilon_{[I, s_a, s_b]}u_n\mid n \in \mathbb{N}_0\} \tag{10.25}
\]

is an orthonormal basis of \(P_{[I, \epsilon_\alpha, \epsilon_\beta, \epsilon_\gamma]}L_2(\mathbb{R})\).

The following theorem present a basis in \(\bigoplus_k P_{[I, \epsilon_\alpha, \epsilon_\beta, \epsilon_\gamma]}L_2(\mathbb{R})\).

**Theorem 10.2.7** Let \(\{u_{n,k}\mid n \in \mathbb{N}_0\}\) be an orthonormal basis in \(L_2(a_k, a_{k+1})\), \(k \in \mathbb{Z}\). Then the system

\[
\{w(t;I_k, \epsilon_k)\epsilon_{[I, s_{k+1}]}u_{n,k} \mid t - a_k \mid n \in \mathbb{N}_0, k \in \mathbb{Z}\} \tag{10.26}
\]

is an orthonormal basis of \(L_2(\mathbb{R})\), where \(s_{k+1}(1) = -s_k(2), \epsilon_k = (\epsilon_{a_k}, \epsilon_{a_{k+1}})\).
For the continuous-time case the famous trigonometric bases are described in the following theorem.

**Theorem 10.2.8** Let \( I = [a, b] \) be a continuous-time interval. In the following we present bases with their corresponding local subspaces in \( L_2(\mathbb{R}) \):

1. The system \( \{w[I, \epsilon_{ab} \sqrt{\frac{2}{T}} \sin \left( \frac{\pi}{T} n (t - a) \right) \mid n \in \mathbb{N}_0 \} \) constitutes a local orthonormal basis in \( P_{[I, a, b], [-1, 1]} L_2(\mathbb{R}) \).

2. The system \( \{w[I, \epsilon_{ab} \sqrt{\frac{2}{T}} \sin \left( \frac{\pi}{T} n (t - a) \right) \mid n \in \mathbb{N}_0 \} \) constitutes a local orthonormal basis in \( P_{[I, a, b], [-1, 1]} L_2(\mathbb{R}) \).

3. The system \( \{w[I, \epsilon_{ab} \sqrt{\frac{2}{T}} \cos \left( \frac{\pi}{T} n (t - a) \right) \mid n \in \mathbb{N}_0 \} \) constitutes a local orthonormal basis in \( P_{[I, a, b], [-1, 1]} L_2(\mathbb{R}) \).

4. The system \( \{w[I, \epsilon_{ab} \sqrt{\frac{2}{T}} \cos \left( \frac{\pi}{T} n (t - a) \right) \mid n \in \mathbb{N}_0 \} \) constitutes a local orthonormal basis in \( P_{[I, a, b], [-1, 1]} L_2(\mathbb{R}) \).

Of course, each frame in \( L_2(I) \) yields a frame in \( P_{[I, a, b], [-\pi, \pi]} L_2(\mathbb{R}) \).

**Theorem 10.2.9** Let \( \{\phi_n \mid n \in \mathbb{N}_0 \} \) be a frame in \( L_2(I) \), with frame bounds \( A, B \) and dual frame \( \{\psi_n \mid n \in \mathbb{N}_0 \} \). Then the system

\[
\{w[I, \epsilon_{ab}] \mathcal{E}_{[I, a, b]} \phi_n \mid n \in \mathbb{N}_0 \}
\]

is a lapped frame in \( P_{[I, a, b], [-\pi, \pi]} L_2(\mathbb{R}) \) with the same frame bounds \( A, B \) and dual frame \( \{w[I, \epsilon_{ab}] \mathcal{E}_{[I, a, b]} \psi_n \mid n \in \mathbb{N}_0 \} \).

The multi-resolution property of the local orthogonal projections, allows for arbitrary refinement in the time domain constructing smooth orthogonal bases. In the next example we represent how the well-known Lemarié-Meyer smooth wavelets can be constructed by the local orthogonal projections.

**Remark 10.2.10** Local orthogonal trigonometric bases allow to construct arbitrarily smooth wavelets. The Lemarié-Meyer smooth wavelets can be obtained by a segmentation of \( (0, \infty) \) in \( \bigcup_{k \in \mathbb{Z}} I_k \), \( I_k = [2^{-k} \pi, 2^{-k+1} \pi] \), \( \epsilon_k = (2^{-k} \pi / 3, 2^{-k+1} \pi / 3) \), \( \delta_k = (+1, -1) \), and the cosine basis, third basis in Theorem 10.2.8. It can be proved that the functions

\[
\tilde{\phi}_{n,k}(\omega) = 2^{\frac{k}{2}} e^{-j2^k n \omega} \tilde{\phi}(2^k \omega), \quad n, k \in \mathbb{Z},
\]
form an orthonormal basis for $L_2(\mathbb{R})$, where the function $\hat{\phi}$ is given by

$$\hat{\phi}(\omega) = \frac{\text{sgn}(\omega) e^{j\frac{\omega}{2}} w(|\omega|; I_0, \frac{\pi}{3}, \frac{2\pi}{3})}{\sqrt{2\pi}},$$

where support($\hat{\phi}$) = $[-\frac{8\pi}{3}, -\frac{2\pi}{3}] \cup [\frac{2\pi}{3}, \frac{8\pi}{3}]$. Applying the inverse Fourier transformation to this basis, we obtain the Lemarié-Meyer smooth wavelets, where the mother generating wavelet is given by

$$\phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{j\omega t} \hat{\phi}(\omega) d\omega,$$

and the elements of the basis in $L_2(\mathbb{R})$

$$\phi_{n,k}(t) = 2^{-\frac{k}{2}} \phi(2^{-k}t - n), \ n, k \in \mathbb{Z}.$$

Other wavelets can be obtained by choosing other dilation, which means changing the support of $\hat{\phi}$. For more details we refer to the literature [3]. □

### 10.3 Best basis selection

In general, there is no straightforward approach to handle the question how the time domain can be segmented. However, in practice, different techniques based on rate distortion are dealing with this question. We refer to [28], for instance, where a spectral entropy is defined to determine the boundaries of the optimal interval segmentation: Let $\mathcal{B} = \{\mathcal{B}[\theta] | \theta \in \Theta \}$ be a ‘library’ collection of orthonormal bases in a segmented Hilbert signal space $\mathcal{H}$, where each basis in the library corresponds with a segmentation of the Hilbert space $\mathcal{H}$. For instance $\mathcal{H} = \ell_2(I)$ with $I$ is a discrete interval, $\#I = L$. The interval $I$ is divided into sub-intervals $I_k$, $k \in \Gamma$. The collection $\mathcal{B}[\theta; k] = \{\phi_n[\theta; k] | n = 0, \ldots, L_k - 1 \}$ is an orthonormal basis in $\mathcal{P}_{[I_k]} \mathcal{H}$, where the $\mathcal{P}_{[I_k]}$’s are orthogonal projections splitting $\mathcal{H}$ as a direct sum. The collection $\oplus_{k \in \Gamma} I_k$ is a segmentation of the underlying time domain of $\mathcal{H}$ and $(\theta; k)$ denotes the boundaries of the intervals $I_k$. Here the parameter $\theta$, as in Chapter 3, can be used for optimization. Define the spectral entropy

$$C(\mathcal{B}[\theta], x) := \sum_{n,k} C\left(\frac{|\langle x, \phi_n[\theta; k] \rangle|^2}{\|x\|^2}\right), \quad (10.28)$$

where $C$ is a concave function, for example $C(x) = -x \log x$. The spectral entropy measures the rate of decrease of the coefficients in the expansion of a signal, after rearranging into decreasing order [99]. In [64] it is proved that if $C(\mathcal{B}[\theta_1], x) < C(\mathcal{B}[\theta_2], x)$, then for all $N \leq L$, the basis $\mathcal{B}[\theta_1]$ concentrate the energy of $x$ in $N$.
largest terms better than any $N$ elements of the basis $B[\theta_2]$. The optimal segmentation corresponding to the optimal $\hat{\theta}$ for $B$ is obtained by

$$C(B[\hat{\theta}], x) := \min_{\theta \in \Theta} C(B[\theta], x),$$

(10.29)

deepends on the choice of the signal $x$. The segmentation for $x$ is in general determined recursively, using dyadic trees, and wavelet packets or local trigonometric bases. Assuming that the library $B$ is a binary tree [28], that is:

1. Subsets of bases can be identified with intervals of the form

$$I_{r,s} = [2^r s, 2^r (s + 1)),$$

for $r, s \geq 0$.

2. Each basis in $B$, $B[r, s]$, corresponds to a disjoint cover of $2^r$ by intervals $I_{r,s}$.

3. For each subspace $\mathcal{V}_{r,s}(= \mathcal{P}_{I_{r,s}} \mathcal{H})$ identified with $I_{r,s}$, we have

$$\mathcal{V}_{r+1,s} = \mathcal{V}_{r,2s} \oplus \mathcal{V}_{r,2s+1}.$$ 

Then the best basis can be determined by induction on $r$: Denote the basis identified with $I_{r,s}$ by $B[r, s]$, then $B[r, s]$ is a basis for $\mathcal{V}_{r,s}$, and denote by $\hat{B}[r, s; x]$ the best basis for a signal $x$ in $\mathcal{V}_{r,s}$. For $r = 0$, there is a single basis corresponding to $I_{0,s}$, which is therefore the best basis: $\hat{B}[0, s; x] = B[0, s], s \geq 0$. For all $r \geq 1$, the best basis $\hat{B}[r, s; x]$ for $x$ can be recursively constructed by:

$$\hat{B}[r, s; x] = \begin{cases} B[r, s; x] & \text{if } C(B[r, s], x) < \\ C(\hat{B}[r - 1, 2s; x], x) + C(\hat{B}[r - 1, 2s + 1; x], x), & \text{otherwise.} \end{cases}$$

(10.30)

**Example 10.3.1** To illustrate the method above, we consider the following example. Let $x$ be a signal of 4 samples, say $x$ is defined on the discrete interval $I = [0, 3]$. The interval $I$ leads to a binary tree of two levels:

$$I_{r,s} = [2^r s, 2^r (s + 1)), r = 0, 1, 2; \ s = 0, \ldots, 2^2 - r - 1,$$

where $I_{2,0} = [0, 4) = I$. The intervals $I_{r,s}$ are placed inside the nodes in the binary tree in the left plot of Fig. 10.3. In the right plot of Fig. 10.3 we present the corresponding fictive cost-tree. This means, for example, for $I_{1,0}$ we have the cost...
Now we apply the recursive algorithm above: we start by marking all the bottom nodes, since $B[0, s] = B[0, s]$. Recursively we try to reduce the cost from the bottom to the top. If a parent node is of lower cost than its childs, i.e. $C(B[r + 1, s], x) < C(B[r, 2s; x], x) + C(B[r, 2s + 1; x], x)$, then we mark the parent. Otherwise we do not mark the parent, but we assign to it the lower cost of its childs. The recursive algorithm results in the tree in the left part of Fig. 10.4, which corresponds with the marked intervals, $I_{0,0}, I_{0,1}, I_{0,2}, I_{0,3}$. We note that the algorithm terminates since the length of $I$ is finite (that is, finite number of levels), and since each node is considered at most twice, as child and as parent.

To determine the best segmentation, we start from the top and take the topmost marked nodes. The remaining nodes in the subtrees of the topmost marked nodes are pruned away. The optimal segmentation in this case is presented in the right part of Fig. 10.4, which is $I_{0,0} \oplus I_{0,1} \oplus I_{1,1}$. The corresponding optimal basis is $B[0, 0] \oplus B[0, 1] \oplus B[1, 1]$. In terms of $\theta$, the optimal parameter is $\hat{\theta} = (0, 0, 1, 1, 2, 3)$. Finally, in terms of $\mathcal{V}$, the optimal segmentation $\hat{\mathcal{V}}(x)$ of $\mathcal{H}$ for $x$:

$$\hat{\mathcal{V}}(x) = \mathcal{V}_{0,0} \oplus \mathcal{V}_{0,1} \oplus \mathcal{V}_{1,1}.$$
Given a segmentation \( \bigcup_{k \in \Gamma} I_k \), \( \Gamma \subseteq \mathbb{Z} \), and a basis \( \mathcal{B}[\theta] = \{ \phi_n[\theta] | n = 1, \ldots, L \} \) for \( x \), it is desired for signal estimation and data compression to approximate the signal \( x \) in a few number of basis elements, say \( N \), not necessarily the first \( N \) basis elements of \( \mathcal{B}[\theta] \):

\[
x_N = \sum_{i=0}^{N-1} \langle x, \phi_{n_i}[\theta] \rangle \phi_{n_i}[\theta],
\]

(10.31)

where \( J_N = \{ n_i | i = 0, \ldots, N - 1 \} \) is chosen according a certain selection criterion, for instance extracting the coefficients containing the most or a certain percentage of the energy of \( x \), that is, the largest inner product amplitude \( \langle x, \phi_n[\theta] \rangle \). In other words, minimizing the quadratic error:

\[
\mathcal{E}_{RT}(N, J_N) := ||x - x_N||^2 = \sum_{n \notin J_N} |\langle x, \phi_n[\theta] \rangle|^2.
\]

(10.32)

### 10.4 Discussion

We have considered the basic construction of local decompositions, by lapped orthogonal projections, for the one-dimensional discrete and continuous-time signals. It is clear that local lapped decompositions with smooth transitions are needed for the reduction of the segmentation effects, as we have seen in the examples. Furthermore, lapped orthogonal decompositions give in general more accurate approximations than the ones obtained by the characteristic function. Even for functions without a harmonic character, for example \( x(t) = t \).

Time signals can be analyzed by lapped orthogonal segmentation as follows. First we apply a pre-segmentation for the given signal. This is a simple segmentation that determines the quantity of the information where we apply the techniques to. It is too expansive to treat directly a segment of duration 10 sec sampled at a rate of 44.1 Hz. For each one of these segments, we apply now the best basis selection algorithm to segment the time-interval of the segment according to the chosen basis, for example, splitting it into quasi-stationary parts.

The systems used in this representation yield a simple implementation. Their use in the lapped orthogonal projections, gives a simple form for the reconstruction. Introducing redundancy by local frames we can find more accurate approximations. Since we introduce a degree of freedom that permits us to add extra frequencies in the local orthogonal bases. This may approximate the peaks in the local spectra more accurately.

This kind of technique appears to be an attractive alternative for existing methods in sinusoidal coding of audio signals [39, 45, 47, 58, 72, 93].
Orthogonal transformations by 1-D domain segmentations
Chapter 11

Orthogonal transformations by 2-D domain segmentations

Local signal analysis in both audio and image coding represents the properties of signals in selected bounded domains. This type of analysis can be carried out using local bases chosen such that only imperceptible blocking effects occur.

In the one-dimensional case, the time axis can be decomposed into partly overlapping intervals. Each of these intervals is associated with an orthogonal projection, in which the overlap allows for some smoothness at the interval boundaries. In this way, we arrive at a set of mutually orthogonal projections. Thus, the one-dimensional signal space can be decomposed into a direct sum of mutually orthogonal subspaces that are images of the signal space under the orthogonal projection associated with the interval decomposition of the real line. For each of these subspaces, we can find bases such that a local orthogonal decomposition of signals can be achieved.

The one-dimensional case can be translated to the two-dimensional case, replacing the interval decomposition by a decomposition in polytopes. But the two-dimensional case introduces additional problems. Now vertices appear where more than one overlap occurs.

A natural approach is replacing the 1-D intervals by the 2-D rectangles segmenting the two-dimensional space. Herewith, choosing the rectangles as the polytope segmentation. Also in this chapter, we show that there is a possibility to choose a segmentation into hexagons. This segmentation also leads to a system of mutually orthogonal projections, which decomposes the 2-D space in a smooth way. In this chapter, we consider 2-D signals in $\ell_2(\mathbb{Z}^2)$. For the continuous case, we refer to the papers [79, 80].

11.1 Two-dimensional domain segmentation

In the one-dimensional case, there is only one type of polytope, namely intervals. In the two-dimensional case, we have many ways to decompose $\mathbb{Z}^2$ into overlapping polytopes. We can use, for example, triangles, rectangles, hexagons, or octagons. In
principle, the space can be divided using a combination of all these types, depending on the analysis we want to apply.

If a subdivision is chosen, we then need to find a decomposition of the signal space by choosing the projections associated with the subdivision. Finally, we need to find a representation for each subspace of the decomposition, in order to accomplish our goal.

We consider the discrete two-dimensional plane $\mathbb{Z}^2$ and the signal vector space $L_2(\mathbb{Z}^2)$. Let $P_{(z_0^+)^2, e^+, s^+}$ be an orthogonal projection on the signal space associated with the discrete half-plane defined by $\{ t = (t_1, t_2) \in \mathbb{Z}^2, t_1 \geq 0 \}$, $e^+ \geq 0$, having the same properties as $P_{(z_0^+)^2, e^+, s^+}$ in the one-dimensional case, that is:

$$
\left( P_{(z_0^+)^2, e^+, s^+} \right) (t) = w^2(t; e^+) x(t) + s^+ w(t; e^+) w(St; e^+) x(St),
$$

for $t \in \mathbb{Z}^2$, where

$$
St = (-1 - t_1, t_2).
$$

The window $w[e^+]$ is a non-negative sequence with $w(t; e^+) = 1$ for $t_1 > e^+$, satisfying the ‘power complementarity condition’:

$$
w^2(t; e^+) + w^2(St; e^+) = 1,
$$

where $\text{support}(w[e^+]) = \{ t \in \mathbb{Z}^2 | t_1 \geq -e^+ \}$. $S$ is a symmetry with respect to the line $\{ t \in \mathbb{R}^2 | t_1 = -\frac{1}{2} \}$, regarding $\mathbb{Z}^2$ as a part of $\mathbb{R}^2$.

The orthogonal projection $P_{(z_0^+)^2, e^+, s^+}$ can be seen as the multiplication with the characteristic function on the half-plane $\{ t \in \mathbb{Z}^2 | t_1 > e^+ \}$, and, on the strip $\{ t \in \mathbb{Z}^2 | -e^+ \leq t_1 \leq e^+ \}$ a correction is applied to achieve some smoothness. Furthermore, note that we obtain two projections, because of the freedom of the sign of $s^+$. Later on, this freedom will be used to achieve mutual orthogonality of the projections. For the half-plane $\{ t \in \mathbb{Z}^2 | t_1 < 0 \}$ we can construct orthogonal projections in two ways. The first one uses the same idea as above, but now replacing the window $w(t; e^+)$ for the right half-plane by $w(St; e^+)$ for the left half-plane. The second way is to use a rotation, with angle of rotation $\pi$. In fact, in this way we can construct a projection on any half-plane of $\mathbb{Z}^2$. Indeed, let $R_\alpha$ be a rotation over an angle $\alpha$ and define the projection $P_{(z_0^+)^2, e^+, s^+}$ by:

$$
P_{(z_0^+)^2, e^+, s^+} = R_\alpha P_{(z_0^+)^2, e^+, s^+} R_{-\alpha},
$$

where the rotation $R_\alpha$ is defined on $L_2(\mathbb{Z}^2)$ by

$$
(R_\alpha x) (t) = x(R_\alpha t), \ t \in \mathbb{Z}^2.
$$
It is easy to check that:

\[(P_{\alpha\beta}^d)_{t_1 = t}^\beta = w^2(R_{\alpha\beta}^d)_{t_1}^\beta \dot{x}(t) + s^+w(R_{\alpha\beta}^d)_{t_1}^\beta u(SR_{\alpha\beta}^d)_{t_1}^\beta x(S_R^d)_{t_1}^\beta\]

where

\[S_\alpha := R_{-\alpha}S_{\alpha},\]

which is an orthogonal projection associated with the half-plane obtained by rotating the half-plane \(\{ t \in Z^2 | t_1 \geq 0 \}\) with angle \(\alpha\), that is, \(\{ t \in Z^2 | t \cos \alpha - t_2 \sin \alpha \}\).

By applying translations, \((T_x)(t) = x(t + d)\), we can define new orthogonal projections associated with any half-plane of \(Z^2\), which have the same properties as the existing projections:

\[P_{\alpha\beta}^d = T_{-d}P_{\alpha\beta}^d T^d\]

where \(T_d x(t) = x(t + d)\) and \(d \in Z^2\). Using these projections \(P_{\alpha\beta}^d\) we would like to introduce orthogonal projections which are associated with polytopes in \(Z^2\). Observe that for \(d = (\beta, 0)\),

\[P_{\alpha\beta}^d = P_{\alpha\beta}^d\]

if \(\beta = \epsilon + \epsilon^+\). This means that \(P_{\alpha\beta}^d\) is an orthogonal projection associated with the strip of the plane \(Z^2\) defined by \(\{ t \in Z^2 | -\epsilon^+ \leq t_1 \leq \beta + \epsilon \}\).

A polytope is an intersection of a finite number of half-planes. To each half-plane there is an associated orthogonal projection \(P_{\alpha\beta}^d\). Taking the product of these projections we obtain an operator which maps a signal in \(\ell_2(Z^2)\) onto a signal with support equal to the polytope. If this product is an orthogonal projection, then the factors in this product should mutually commute. We now return to the two projections \(P_{\alpha\beta}^d\) and \(P_{\alpha\beta}^d\): see (11.1). We may now pose the question: when is \(P_{\alpha\beta}^d\) a projection? Or, equivalently, when does the equality

\[P_{\alpha\beta}^d = P_{\alpha\beta}^d\]

hold? The expressions of \(P_{\alpha\beta}^d\) and \(P_{\alpha\beta}^d\) are
explicitly given by:

$$\left( P_{[z \alpha]^2}^{\epsilon+} s^{+1} \right) (t)$$

$$= w^2 (t; \epsilon^+) w^2 (R \alpha t; \epsilon) x(t)$$

$$+ s^+ w(t; \epsilon^+) w(S t; \epsilon^+) w^2 (R \alpha t; \epsilon) x(S t)$$

$$+ s w(R \alpha t; \epsilon) w(S R \alpha t; \epsilon) w^2 (t; \epsilon^+) x(S \alpha t)$$

$$+ s s^+ w(t; \epsilon^+) w(S t; \epsilon^+) w(R \alpha t; \epsilon) w(S R \alpha t; \epsilon) x(S \alpha S t).$$

The necessary and sufficient condition for the equality in (11.2) to hold, is that the symmetries applied in the projections $P_{[z \alpha]^2}^{\epsilon+}$ and $P_{[z \alpha]^2}^{\epsilon+}$ commute, that is:

$$S \alpha S = S \alpha.$$

Consequently, the symmetries used here must form a commutative group, namely the group $\{L, S, S \alpha, SS \alpha\}$, cf. [10]. This is satisfied if $\alpha = k \pi/2, k \in \mathbb{Z}$. If we want to proceed this way, the segmentation has to be in rectangles. So for example, a subdivision into triangles will not satisfy this requirement. If we want to use the projections introduced in this work, the subdivision requires some symmetric pattern. In fact, we want to use one basic unit or object, to build a symmetric pattern in a unique way. This idea corresponds to using a Dirichlet domain, namely the smallest area enclosed by the bisection lines in a given lattice [46, 83, 84]. In total, there are five domains in the two-dimensional case, two rectangular (square and non-square) and three hexagonal forms.

In the rectangular case, where the symmetries commute, we can define an orthogonal projection associated with a rectangle $\square$ as

$$P_{\square} = P_{[z \alpha]^2}^{d\frac{\pi}{2} d\alpha} P_{[z \alpha]^2}^{d\frac{\pi}{2} d\alpha} P_{[z \alpha]^2}^{d\frac{\pi}{2} d\alpha} P_{[z \alpha]^2}^{d\frac{\pi}{2} d\alpha} P_{[z \alpha]^2}^{d\frac{\pi}{2} d\alpha}.$$
for a suitable choice of of the vectors $d_{11}, d_{12}, d_{21}$ and $d_{22}$, characterizing the place and the size of the rectangle; see Fig. 11.1. To accomplish the decomposition of the signal space $\ell_2(\mathbb{Z}^2)$:

$$
\ell_2(\mathbb{Z}^2) = \bigoplus_k P_{\square_k} \left( \ell_2(\mathbb{Z}^2) \right),
$$

where $\{\square_k|k\}$ is a rectangular subdivision of $\mathbb{Z}^2$, we have to require that $P_{\square_k} P_{\square_l} = \mathcal{O}$, for $k \neq l$, which is achieved by a systematic choice of the overlaps $\epsilon(k)$ and the polarities $s(k)$:

$$
\epsilon(k) = \begin{pmatrix}
\epsilon_{11}^{(k)} & \epsilon_{12}^{(k)} \\
\epsilon_{21}^{(k)} & \epsilon_{22}^{(k)}
\end{pmatrix},
\quad
s(k) = \begin{pmatrix}
s_{11}^{(k)} & s_{12}^{(k)} \\
\tilde{s}_{11}^{(k)} & \tilde{s}_{12}^{(k)}
\end{pmatrix}.
$$

In each of the spaces $P_{\square_k} \left( \ell_2(\mathbb{Z}^2) \right)$ we can choose a suitable signal representation for a local decomposition. Thus, the goal to find a local signal representation in $\ell_2(\mathbb{Z}^2)$ is achieved. It is even possible to construct a basis in $P_{\square_k} \left( \ell_2(\mathbb{Z}^2) \right)$ by defining the symmetrical extension for 2-D signals; see (10.8) in previous chapter. Using this symmetrical extension for 2-D signals and an orthonormal basis for $\chi_{\square} \left( \ell_2(\mathbb{Z}^2) \right)$, an orthonormal basis for $P_{\square_k} \left( \ell_2(\mathbb{Z}^2) \right)$ is obtained, in a similar way as in the one-dimensional case. A different approach is presented in [10]. In [11], the problem of window design, especially for rectangular segmentations, is discussed. In this paper the authors mentioned that the techniques developed for the rectangular segmentation can be directly used for hexagonal segmentation. In the next section we show that some manipulations are needed to solve the hexagonal problem.

## 11.2 Subdivision into hexagons

We now consider a hexagonal subdivision, shown in Fig. 11.2, with in the hexagon two angles $\pi/2$ and four angles $3\pi/4$. For simplicity, we will choose a subdivision $\{H_{i,j}|i,j \in \mathbb{Z}\}$ where each $H_{i,j}$ is a shifted version of a fixed prototype hexagon $H$ (with $H_{0,0} = H$). To make this discussion clear we will fix $(i,j) \in \mathbb{Z}^2$ and
concentrate on one hexagon $H_{i,j}$ which we then rotate over an angle $\pi/4$, in order to obtain $H'_{i,j}$: see Fig. 11.3. In fact, the rotation over $\pi/4$ has nothing to do with the segmentation, because any angle can be taken. We have taken this specific angle to get an orientation where we can describe our ideas. So if we have determined the orthogonal projection $\mathcal{P}_{i,j}$ associated with $H_{i,j}$, then $\mathcal{P}_{i,j} = R_{-\pi/4} \mathcal{P}_{i,j} R_{\pi/4}$ is the orthogonal projection associated with $H'_{i,j}$.

If the orthogonal projection $\mathcal{P}_{i,j}$ is constructed, then the construction allows to obtain all other projections by shifting $\mathcal{P}_{i,j}$. Therefore, it makes sense to concentrate on a single $H_{i,j}$. From Fig. 11.3, we can see that each hexagon of the subdivision can be partitioned into thirteen different and disjoint zones. Such zones are denoted by $C_{i,j}^k$, where $k = 0, 1, \ldots, 12$. We note that $C_{i,j}^0$ is a hexagon, a scaled version of $H_{i,j}$.

The next step is to decompose the signal space into subspaces associated with these zones. We note that if $\mathcal{P}^{C_{i,j}^k} = \chi_{C_{i,j}^k} \mathcal{P}_{C_{i,j}^k}$ are orthogonal projections associated with $C_{i,j}^k$, then

$$
\mathcal{P}_{i,j} = \sum_{k=0}^{12} \mathcal{P}^{C_{i,j}^k}
$$

is an orthogonal projection associated with $H_{i,j}$; with support($\mathcal{P}_{i,j}x$) = $H_{i,j}$.

The central zone $C_{i,j}^0$ has no intersection with other zones, therefore the projection $\mathcal{P}^{C_{i,j}^0}$ must be $\chi_{C_{i,j}^0}$. The zone $C_{i,j}^1$ corresponds to two zones $C_{i-1,j}^0$ and $C_{i-1,j+1}^0$ of $H_{i-1,j}$ and $H_{i-1,j+1}$, respectively. Therefore, $\mathcal{P}^{C_{i,j}^1} + \mathcal{P}^{C_{i-1,j}^0} + \mathcal{P}^{C_{i-1,j+1}^0} = \chi_{C_{i,j}^1}$ must hold. If we apply these projections to the signal space $\ell_2(\mathbb{Z}^2)$, it means that the direct sum of the three vector spaces associated with $C_{i,j}^0$, $C_{i-1,j}^0$ and $C_{i-1,j+1}^0$ yields the signal subspace $\chi_{C_{i,j}^1} (\ell_2(\mathbb{Z}^2))$. The zones $C_{i-1,j}^0$ and $C_{i-1,j+1}^0$ have the same structure as $C_{i,j}^0$. So it is sufficient if we concentrate on the zones of $H_{i,j}$. From the...
### 11.2 Subdivision into hexagons

**Figure 11.3: Decomposition of a hexagon $\tilde{H}_{i,j}$ into the overlapping sets $C_{i,j}$.**

In the rectangular case, we write for the orthogonal projection for $C_{i,j}^{1}$:

$$
\left( P_{C_{i,j}^{1}} x \right)(t) = w_{i,j}(t) \left[ w_{i,j}(2a - 1 - t_1, t_2) + s_1 w_{i,j}(2a - 1 - t_1, t_2) + s_2 w_{i,j}(2a - 1 - t_1, 2a - 1 - t_2) + s_1 s_2 w_{i,j}(2a - 1 - t_1, 2a - 1 - t_2) \right].
$$

by considering $C_{i,j}^{1} = [0, 2a - 1] \times [0, 2a - 1]$. $w_{i,j}$ has support $C_{i,j}^{1}$ and satisfies:

$$
\begin{align*}
& w_{i,j}^{2}(t_1, t_2) + w_{i,j}^{2}(2a - 1 - t_1, t_2) + w_{i,j}^{2}(t_1, 2a - 1 - t_2) + \\
& w_{i,j}^{2}(2a - 1 - t_1, 2a - 1 - t_2) = 1
\end{align*}
$$

and $s_1, s_2 \in \{-1, 1\}$, where we have chosen $\epsilon_1 = \epsilon_2 = a$. An example of a window satisfying the above condition is $w_{i,j}(t_1, t_2) = g(t_1)g(t_2)$ where

$$
g(\tau) = \sin\left( \frac{\pi(\tau + 1)}{4a + 2} + \frac{1}{4} \cos \frac{2\pi(\tau + 1)}{2a + 1} - \frac{1}{4} \right) \text{ for } \tau \in [0, 2a - 1].
$$

Note that the projection for $C_{i,j}^{1}$ can be constructed in the same way.

Summarizing the results above, we note the following:

- we have a zone as an intersection of three hexagons, namely $C_{i,j}^{1}$, $C_{i-1,j}^{0}$ and $C_{i-1,j+1}^{0}$, and,

- we have four projections with the structure of $P_{C_{i,j}^{0}}$, which are orthogonal and satisfy the complementarity condition, that is, the sum of these projections
equals $\chi_{G_1^{ij}}$. The projections can be obtained by use of the four symmetries 
\{$(t_1, t_2), (2a - 1 - t_1, t_2), (t_1, 2a - 1 - t_2), (2a - 1 - t_1, 2a - 1 - t_2)$\}, and 
by a suitable choice of the polarities $s_1$ and $s_2$.

Secondly, the necessary condition acting on the windows of the projections associated 
with $C_{i-1,j}^0$ and $C_{i-1,j+1}^5$ is that they must vanish on just one side instead of two. 
Referring to Fig. 11.1 the window corresponding to $C_{i,j}^1$ must vanish at the sides $d_{11}$ 
and $d_{21}$. The window corresponding to $C_{i-1,j}^0$ must vanish at $d_{22}$, and the window 
corresponding to $C_{i,j+1}^5$ must vanish at $d_{12}$.

The zone $C_{i,j}^2$ only intersects the zone $C_{i,j}^6$. To have the orthogonal projection, 
then $P_{C_{i,j}^2} + P_{C_{i,j}^6} = \chi_{C_{i,j}^2}$ must hold, and, of course, $P_{C_{i,j}^2} P_{C_{i,j}^6} = \mathcal{O}$, but we 
also require that $P_{C_{i,j}^1} + P_{C_{i,j}^2}$ yields a ‘smooth’ window. This can be achieved by 
using the window \( w_{i,j}(t_1, t_2) = g(t_2) \) for \( (t_1, t_2) \in [2a, b+2a] \times [0, 2a-1] \) if $C_{i,j}^2$ is 
represented by $C_{i,j}^2 = [2a, 2a+b] \times [0, 2a-1]$. The same arguments can be applied 
to $C_{i,j}^3$ and $C_{i,j}^4$.

Denoting the projection associated with the zone $C_{i,j}^3$ by $Q_{C_{i,j}^3}$, then for its con-
struction we shall use the projection $P_{C_{i,j}^1}$, that is,

\[
Q_{C_{i,j}^3} = \mathcal{T}^{-2a-b} \mathcal{R}_{\pi/2} P_{C_{i,j}^1} \mathcal{R}_{-\pi/2} \mathcal{T}^{2a+b}.
\]

The zone $C_{i,j}^4$, which corresponds to $C_{i,j-1}^{10}$, has a non-rectangular form. We can 
construct the projection here by extending the window of $C_{i,j}^3$ (the values of the 
window at the two common sides with $C_{i,j}^3$) on lines parallel to the oblique sides of 
$C_{i,j}^3$; see Fig. 11.3. In this way, we obtain an orthogonal projection with a window 
which is zero in a half-zone of $C_{i,j}^4$. Consequently, we have two new problems to 
solve:

- we have three projections associated with $C_{i,j}^3, C_{i-1,j}^7$ and $C_{i,j-1}^{11}$ whose sum 
is not equal to the identity, and

- the sum of the two projections of $C_{i,j}^4$ and $C_{i,j-1}^{10}$, obtained in the way as des-
cribed above, is also not equal to the identity.

To solve this problem, we add to $Q_{C_{i,j}^3}$ the orthogonal projection obtained by shifting 
$P_{C_{i,j}^1}$ with $(2a+b, 0)$, so $P_{C_{i,j}^2} = Q_{C_{i,j}^3} + \mathcal{T}^{-2a-b} P_{C_{i,j}^1} \mathcal{T}^{2a+b}$, and extend it, in 
the same way as described above, for $C_{i,j}^2$. We only do this for one side, namely, 
$C_{i,j}^3, C_{i,j}^4$ and $C_{i,j}^5$, and not for $C_{i,j}^0, C_{i,j}^{10}$ and $C_{i,j}^{11}$. For the zones $C_{i,j}^0$,
$C_{i,j}^{10}$ and $C_{i,j}^{11}$ we only use a rotated and shifted version of $P_{C_{i,j}^1}$. For the zone $C_{i,j}^5$, the orthogonal 
projection is a rotated and shifted version of $P_{C_{i,j}^1}$. In this way, the two problems
are solved, and, therefore, the projection associated with $\hat{H}_{i,j}$. For $H_{i,j}$ we have, $\mathcal{P}_{i,j} = \mathcal{R}_{-\pi/4} \hat{\mathcal{P}}_{i,j} \mathcal{R}_{\pi/4}$.

For a basis for the local space $\mathcal{P}_{i,j} (\ell_2(\mathbb{Z}^2))$, a similar result by symmetrical extension, as for rectangles can be derived for hexagons. The main question is then: Find a basis with a hexagon as its overall support? This is a tough question. In practice, one is more interested in representations, that describe the signals completely, and not necessarily representations based on orthonormal bases. Moreover, it is sometimes better to use redundant representations. Theorem 10.1.12 tells us how simple it can be to construct tight frames for a subspace of signals; with a polytope as support. This theorem is formulated in the general context of an abstract Hilbert space. So it can be applied for finding tight frames in the one-dimensional case; see previous chapter where we have shown how to construct tight frames. For the hexagonal segmentation, we can apply the theorem as follows. Given a hexagon $H_{i,j}$, consider a basis with as support the smallest rectangle containing the hexagon, for example, the 2-D cosine basis $\{\psi_n | n\}$. Then the system $\{\mathcal{P}_{i,j} \psi_n | n\}$ constitute a tight frame in $\mathcal{P}_{i,j} (\ell_2(\mathbb{Z}^2))$. Joining all these local tight frames we obtain a localizing tight frame $\{\mathcal{P}_{i,j} \psi_{n,i,j} | n, i, j\}$ in $\ell_2(\mathbb{Z}^2)$ for the hexagonal segmentation.

11.3 Discussion

The one-dimensional solutions can be used for the two-dimensional case, but not without restricting the subdivision of the space. It is shown that in the case of rather arbitrary (convex) subdivisions, additional problems arise that do not exist in the one-dimensional case. In particular, the hexagonal subdivision avoids these problems.

Local total systems with smooth transitions are needed for the reduction of blocking effects as occur in the case of non-smooth transitions in, for example, audio and image coding. The representations yield a simple implementation: their use in the lapped orthogonal transformation gives a simple form of its inverse transformation, namely its transpose, since it is orthogonal.

For subdivisions where we cannot find a commutative group of symmetries (different from the identity), we have to look for other alternatives to accomplish our goal. In [80], we discussed two alternatives, one based on relaxing the requirement of mutual orthogonality of the projections, the other one based on non-linear symmetries.
Orthogonal transformations by 2-D domain segmentations
Bibliography


Samenvatting

Signaalrepresentaties spelen een fundamentele rol in de signaalverwerking, in het bijzonder bij het ontwerpen van signaalverwerkingssystemen. Doelmatigheid en kwaliteit van zo’n systeem hangen sterk af van de gekozen signaalrepresentatie. Vaak zijn nauwkeurige benaderingen en snelle algoritmen vereist bij het representeren en analyseren van signalen. In dit proefschrift worden twee typen signaalrepresentaties beschouwd:

1. geëparameërde globale signaalrepresentaties;
2. geëparameërde lokale signaalrepresentaties.

Globale representaties hebben betrekking op het hele domein, terwijl lokale representaties voortgebracht worden door zogenaamde korte-termijn transformaties. Dus het eerste type beschrijft globale eigenschappen van het signaal en het tweede type de lokale eigenschappen. De parameters kunnen zo gekozen worden dat de doelmatigheid van de representaties wordt vergroot.

In dit proefschrift worden verschillende problemen met betrekking tot doelmatigheid en kwaliteit van signaalrepresentaties geformuleerd en geanalyseerd. Een convergentiecriterium voor welbekende reeksontwikkelingen (zoals Jacobi, Hermite, Laguerre, Kautz) wordt geformuleerd. Op basis van dit criterium worden optimale parameter keuzes voor deze reeksontwikkelingen afgeleid.

Het concept “gevensterde unitaire transformatie” wordt geïntroduceerd; het generaliseert het concept “gevensterde Fourier transformatie”. Er wordt uitvoerig ingegaan op relaties tussen het nieuw-geïntroduceerde concept, gevensterde unitaire transformatie, en het klassieke concept filterbank. Eenvoudig te interpreteren representatieschema’s voor perfect-reconstruerende filterbanken worden ontwikkeld.

Artefacten aan de randen die veroorzaakt worden door de opdeling van het signaal in onafhankelijke signaalsegmenten door gebruik van rechthoekige vensters kunnen worden onderdrukt door deze vensters te vervangen door gladde vensters en door symmetrie-voorwaarden toe te passen aan de randen van het gladde venster. De overeenkomstige lokale signaalrepresentaties worden zowel voor het een-dimensionale als voor het twee-dimensionale signaaldomein beschreven.
Dankwoord

Op deze plaats wil ik al diegenen bedanken die mij gedurende mijn promotieperiode direct danwel indirect ondersteund hebben.

Allereerst wil ik mijn copromotor Bert den Brinker bedanken voor het initiëren van mijn promotie-onderzoek, de vele vruchtbare discussies en de plezierige samenwerking. Tevens wil ik Stef van Eijndhoven bedanken voor zijn toewijding aan mijn promotie-onderzoek. In het bijzonder voor de hechte samenwerking, niet alleen met betrekking tot het onderzoek maar ook op persoonlijk vlak. Bert en Stef, heel hartelijk bedankt voor jullie betrokkenheid, tijd en inzet!


Daarnaast bedank ik mijn kamergenoten Arno van Leest (Arno-san) en Daniël Schobben (Daniël-san). Vriend Arno-san, bedankt voor de prettige vier jaar die wij als kamergenoten hebben doorgebracht. Mijn dank gaat ook uit naar Achée Lin (Achée-san) voor de gezellige tijd en naar Jurgen van Engelen (Jurgen-san) voor het beschikbaar stellen van de \LaTeXX-style van zijn proefschrift, waardoor dit proefschrift deze vorm heeft gekregen. Collegha’s, bedankt voor de XBlast-tijd! Verder bedank ik Thomas de Hoog voor het ter beschikking stellen van historische literatuurreferenties.


Mijn vader en moeder bedank ik voor alle mogelijkheden die ze mij gegeven hebben en hun onvoorwaardelijke steun bij alle keuzes die ik gemaakt heb. Mijn zussen Kautar, Dikra en Sara, en broers, Zakaria, Mouad, Youssef, Ilyas en Issam wil ik bedanken voor de hechte familieband. Tenslotte, bedank ik mijn lieve vrouw Nabila en mijn lieve zoon Ziade voor hun geduld, steun en het bieden van een gelukkig en stabiel thuisfront.

Bahaa Eddine Sarroukh
21 november 2000
Curriculum Vitae

7 Sept. 1971  Born in Larache, Morocco.


1991  Dutch course, Advies- en Studiecentrum, Heerlen, the Netherlands.

1992-1996  Master student at the Eindhoven University of Technology, Department of Mathematics and Computing Science, Systems and Control Group. The master project entitled “Filter banks induced by unitary integral transformations” was carried out under supervision of prof.dr.ir. M.L.J. Hautus, dr.ir. S.J.L. van Eijndhoven and dr.ir. A.C. den Brinker.

**List of Publications**


